University of Alberta

# METHODS IN STATISTICAL SHAPE ANALYSIS FOR LANDMARK-BASED THREE-DIMENSIONAL DATA, WITH APPLICATIONS TO AN ORTHODONTIC STUDY

by

Jennifer Pamela Gamble

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of
Master of Science

in

Statistics

Department of Mathematical and Statistical Sciences
Edmonton, Alberta
Fall, 2008

Library and Archives Canada

Published Heritage Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et Archives Canada

Direction du Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

*Your file Votre référence*
*ISBN: 978-0-494-47251-4*
*Our file Notre référence*
*ISBN: 978-0-494-47251-4*

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

---

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

**University of Alberta**

**Library Release Form**

**Name of Author:** Jennifer Pamela Gamble

**Title of Thesis:** Methods in statistical shape analysis for landmark-based three-dimensional data, with applications to an orthodontic study

**Degree:** Master of Science

**Year This Degree Granted:** 2008

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly, or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

_____

Date:

*To my parents.*

*I am very happy with my life, and I attribute that to you.*

# ABSTRACT

The field of statistical shape analysis is aimed at describing and comparing the shapes of objects. Areas of application are often in fields related to medicine or biology, with shapes of interest that are often complex and non-geometric. Shape analysis methods allow clinicians and researchers to analyze shape differences and shape changes within individuals, across groups, and over time. These can occur due to growth, varying treatment techniques, or natural differences between groups.

This thesis will discuss some methods that have been developed in the field of landmark-based statistical shape analysis, and apply these methods to an example three-dimensional data set taken from the field of orthodontics. Landmark-based methods apply to shapes that have been represented as a set of landmarks, or points, of interest (as opposed to shapes that are represented by boundary information, or an entire pixelated/voxelated image). Shape analysis methods involve transforming these landmark configurations into some shape space (or approximation to it), in order to define distances between shapes, and perform statistical analyses.

Because of some statistical issues that arise in the standard shape analysis setting, an alternative approach will also be applied. This approach, borrowed from computational topology, is called persistent homology. It will be applied to the existing shape analysis framework to detect clusters or subgroups of similarly-shaped objects.

# ACKNOWLEDGEMENTS

My first and foremost acknowledgement definitely has to go to my supervisor, Dr. Giseon Heo. She convinced me to go to graduate school, and introduced me to the field of shape analysis. I'm so happy that she did, and I have never known of a supervisor so encouraging and supportive. She has always gone out of her way for me: to make sure that I could attend conferences, and to make sure I was enjoying my work. She is consistently insightful, both in terms of research and life.

I also have to extend my thanks to Dr. Byron Schmuland, first for agreeing to be my co-supervisor, and of course for all his financial support over the past two years. My acknowledgements go to NSERC (through Dr. Schmuland), the Orthodontic program's McIntyre Memorial Fund (through Dr. Heo), the Faculty of Graduate Studies and Research, and the Graduate Students Association for their travel funding. Manuel LaGravere and Paul Major in the Orthodontics division of the Department of Dentistry supplied the data set used throughout this thesis.

Thanks to Ryan Trelford for all his help with LaTeX and formatting. Dr. Paul Major and Dr. George Peschke provided thoughtful guidance. Finally, I send my love and thanks to Chris, for his patience all these years, and his ability to keep me sane.

# Table of Contents


**1 Introduction** **1**

1.1 Statistical Shape Analysis . . . . . 2

1.1.1 Landmark Configurations and Superimposition . . . . 3

1.1.2 Shape Spaces . . . . . 5

1.1.3 Procrustes Distances . . . . . 7

1.1.4 Procrustes Methods . . . . . 9

1.1.5 Tangent Space Methods . . . . . 14

1.2 Euclidean Distance Matrix Analysis . . . . . 17

1.2.1 Description of shape and form . . . . . 18

1.2.2 Calculating mean form matrix . . . . . 20

1.2.3 Comparison between groups . . . . . 22

1.2.4 Using EDMA to identify influential landmarks . . . . . 26

1.3 Other methods in shape analysis . . . . . 28

**2 Landmark reliability** **31**

2.1 Description of data set . . . . . 31

2.2 Intraclass Correlation Coefficients (ICCs) as a measure of reliability . . . . . 32

2.3 Dry skull data . . . . . 36

**3 Application of Shape Analysis Techniques** **43**

3.1 Comparing groups at baseline . . . . . 44

3.1.1 Dryden and Mardia methods . . . . . 44

3.1.2 EDMA methods . . . . . 49
3.2 Shape change within groups over time . . . . . 51
3.3 Shape differences between groups at times 2, 3, and 4 . . . . . 53
3.4 Longitudinal shape analysis using repeated measures MANOVA 55
3.5 How to proceed . . . . . 56

**4 Finding Influential Landmarks** **59**
4.1 Mean forms within groups across time . . . . . 59
4.2 Mean forms across treatment groups for given time points . . 61
4.3 Repeated Measures MANOVA . . . . . 63
4.4 Conclusions . . . . . 63

**5 Persistent Homology** **64**
5.1 Background and Theory . . . . . 65
5.1.1 History . . . . . 65
5.1.2 Topological spaces . . . . . 65
5.1.3 Filtrations . . . . . 67
5.1.4 Simplicial Complexes . . . . . 69
5.1.5 Homology . . . . . 73
5.1.6 Persistent Homology . . . . . 77
5.2 Persistent homology applied to tangent space data . . . . . 78
5.2.1 Results . . . . . 79

**6 Directions for Future Research** **83**

**Bibliography** **85**

**A Additional table and figures** **90**

**B Computer code** **104**

# Chapter 1

# Introduction

Statistical shape analysis is a field developed primarily within the last 25 years, with the objective of facilitating descriptions and comparisons of the shapes of objects. In this context, *shape* refers to any geometrical information about an object that is invariant to the *similarity transformations* of rotation, translation, and rescaling. It is often desirable to compare groups of shapes for statistically significant differences in their 'mean shape'. To do this, the distance between two shapes must be quantified somehow. Many common areas of application are in the fields of medicine and biology, with digital representations of the objects of interest obtained through medical imaging methods in 2- and 3-dimensions. Many of the tools developed for statistical shape analysis are intended to analyze the shape of objects that are represented as a set of landmarks, usually of biological or clinical significance.

This thesis will discuss some of these mathematical and statistical tools for analyzing shapes. An orthodontic data set involving 3-dimensional craniofacial landmarks will be used throughout. This introductory chapter will outline the theory of some established methods in statistical shape analysis, which will be applied to the data set in Chapters 3 and 4. Chapter 2 will discuss the data set, and the reliability of landmark placement. With the example data set, as is often the case in real-world studies, it is not only of interest to compare treatment groups at one time-point, but longitudinally as well.

Because of some statistical issues that arise when comparing multiple groups in the standard setting, an alternative, more qualitative method will be discussed. This method, called persistent homology, is from computational topology, and as this thesis shows, can be applied to the existing shape analysis framework to detect similarly-shaped subgroups within the larger set of subjects. This will be presented in Chapter 5, and contrasted with the existing longitudinal methods discussed in Chapter 3. Conclusions and directions for future research will finish the text in Chapter 6.

## 1.1 Statistical Shape Analysis

The idea of using mathematical expressions to describe biological shape differences dates back to the landmark 1917 text *On Growth and Form* by D'Arcy Wentworth Thompson [43]. He discussed concepts such as growth-related shape change, shape differences corresponding to function, and the use of deformation grids to map one shape into another.

The mathematics of shape theory began to be developed formally in the late 1970's and 1980's, with Fred Bookstein ([3], [4], [6]) and David Kendall ([26], [27]). Although methods do exist for dealing with other types of data (see Section 1.3 below), this thesis will focus on analyzing shapes which are represented as a set of landmarks (as opposed to entire curves or surfaces, or pixelated/voxelated data). To this end, we will first review some concepts, following the exposition of Dryden and Mardia [15] on Procrustes methods, shape spaces, and tangent space approximations, as pertaining to landmark data. Their book *Statistical Shape Analysis* [15] can be referenced for further details. For some of the methods described below (specifically those of Dryden and Mardia), the mathematics for two-dimensional data use complex numbers to represent the two dimensions, and thus differ both in technique and notation from the three-dimensional methods. Because all of the data sets considered here will be three-dimensional, the descriptions of the methods for $m = 2$ will be omitted (if they are different than the $m = 3$ case), but can be found in

their text.

### 1.1.1 Landmark Configurations and Superimposition

An important concept in statistical shape analysis is that of *homologous landmarks*. These are reference points placed on the image/object under analysis, which are usually biologically relevant and appropriate for the objective of the study. The landmarks must be correspondingly placed on all objects of interest in a consistent way (discussed further in Chapter 2). The use of landmark data is particularly appropriate in fields (such as dentistry and orthodontics) where well-defined landmarks have already been identified. Examples in that field include the pogonion (the most forward-projecting point on the anterior surface of the chin), or the buccal cusp tip of the right pre-molar.

Once each object has been represented as a configuration of landmarks, we would like to quantify the shape of each configuration. This will allow statements about the similarity/difference of shapes or groups of shapes. The groups of shapes under analysis could correspond to things like gender, treatment group, or the same individuals measured at different time points. This analysis of groups of shapes is where the 'statistical' part of shape analysis comes in. When comparing individuals there is no randomness involved (except for that due to measurement error). Whereas given multiple groups, each is taken to be a sample from some population which has a mean shape, with an unknown distribution of shapes about that mean.

The registration of landmarks for each object is done in some given coordinate system, so a set of $k$ landmarks in $m$ dimensions can be represented by a $k \times m$ *configuration matrix*. Since shape information is invariant to similarity transformations, it is necessary that any measure of shape is as well, and thus independent to the original choice of coordinate system. A method called Procrustes superimposition will be performed to optimally rotate, translate, and scale the configurations (to minimize the squared distances between corresponding landmarks) prior to shape comparison.

**Definition 1.1.** If $X_1$ and $X_2$ are $k \times m$ landmark configuration matrices, then $X_1$ and $X_2$ have the *same shape* if

$$X_2 = \beta X_1 \Gamma + \mathbf{1}_k \gamma$$

where

$\beta \rightarrow$ positive scaling constant

$\Gamma \rightarrow$ $m \times m$ special orthogonal rotation matrix, i.e)$\Gamma \in SO(m)$

$\gamma \rightarrow$ $m \times 1$ translation vector

$\mathbf{1}_k \rightarrow$ $k$-length vector of ones

In the simplest case, matching one landmark configuration $X_1$ as closely as possible to a second configuration $X_2$, Procrustes superimposition consists only of finding the estimates $\hat{\beta}$, $\hat{\Gamma}$, and $\hat{\gamma}$ that minimize

$$\|X_2 - \beta X_1 \Gamma - \mathbf{1}_k \gamma^T\|^2$$

where $\|X\| = \{\text{trace}(X^T X)\}^{1/2}$. This distance is the sum of the squared distances between each pair of homologous landmarks, and the minimizing quantities can be computed directly. The method of matching one configuration onto another is called *full ordinary Procrustes analysis* (full OPA). Note that the minimum distance found matching $X_1$ onto $X_2$ does not necessarily equal that found matching $X_2$ onto $X_1$.

The general case becomes more complicated, with configuration matrices $X_1, X_2, \ldots, X_n$ corresponding to $n \geq 2$ subjects. Now translations, rotations and scalings must be performed in such a way that all the configurations are optimally superimposed. In other words, so that the sum of squared norms of

pairwise distances

$$(1.1) \qquad (1/n)\sum_{i=1}^{n}\sum_{j=i+1}^{m}\|(\beta_i X_i\Gamma_i + \mathbf{1}_k\gamma_i^T) - (\beta_j X_j\Gamma_j + \mathbf{1}_k\gamma_j^T)\|^2$$

is minimized over all $\Gamma_i \in SO(m)$, $\beta_i > 0$, and $\gamma_i$. This method is called *full generalized Procrustes analysis* (full GPA).

If there were no additional restrictions imposed on this, the size of each configuration could be scaled to become arbitrarily small. To avoid this, a restriction on the *centroid size* of the mean configuration is imposed:

**Definition 1.2. Centroid size** is defined as

$$S(X) = \|CX\| = \sqrt{\sum_{i=1}^{k}\sum_{j=1}^{m}(X_{ij} - \bar{X}_j)^2},\ X \in \mathbb{R}^{km},$$

where $X_{ij}$ is the $(i,j)$th entry of $X$, $\bar{X}_j = (1/k)\sum_{i=1}^{k} X_{ij}$ is the arithmetic mean of the $j$th dimension, and $C$ is the $k \times k$ centring matrix, whose entries are $\frac{k-1}{k}$ along the diagonal, and $-\frac{1}{k}$ everywhere else.

To insure that not all $\beta_i$ be close to zero, the centroid size of the mean configuration is restricted to be 1. When working with landmarks in 2-dimensions, there exists an explicit solution which minimizes Eq. 1.1. When the dimension is $m \geq 3$ however, the minimization must be performed iteratively. The details for full OPA and full GPA will be presented in Section 1.1.4.

### 1.1.2 Shape Spaces

When removing similarity transformations to obtain a standardized representation of the shape of a configuration, translation and scaling are filtered out first. Translation is removed through pre-multiplication by a centring matrix, with common choices of either

$$(1.2) \qquad C = I_k - \frac{1}{k}\mathbf{1}_k\mathbf{1}_k^T$$

as used in Definition 1.2, or the $k \times (k-1)$ Helmert sub-matrix defined below.

**Definition 1.3.** The $k \times (k-1)$ *Helmert sub-matrix* $H$ is given by

$$H = \begin{bmatrix} \frac{-1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & \cdots & \cdots & 0 \\ \frac{-1}{\sqrt{6}} & \frac{-1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{-1}{\sqrt{(k-1)k}} & \cdots & \cdots & \frac{-1}{\sqrt{(k-1)k}} & \frac{(k-1)}{\sqrt{(k-1)k}} & 0 \\ \frac{-1}{\sqrt{k(k+1)}} & \cdots & \cdots & \cdots & \frac{-1}{\sqrt{k(k+1)}} & \frac{k}{\sqrt{k(k+1)}} \end{bmatrix}.$$

In other words, for $h_j = -j(j+1)^{-1/2}$, the $j$th row of $H$ contains $h_j$ $j$ times, followed by $-jh_j$, and then $k-j-1$ zeros, for $j = 1, \ldots, k-1$.

Write

$$X_H = HX \in \mathbb{R}^{(k-1)m} \backslash \{0\}, \text{ or alternatively } X_C = CX$$

to represent configurations with location removed. Note the relationship

$$H^T H = C, \text{ and so } H^T X_H = CX.$$

Centroid size (see Definition 1.2) is chosen as the divisor for size-standardization. Since $C$ is idempotent (i.e. $C^T C = C$), also note that

$$S(X) = \|CX\| = \sqrt{\text{tr}(X^T C X)} = \sqrt{\text{tr}(X^T H^T H X)} = \|X_H\|$$

Which leads us to the definition of pre-shape:

**Definition 1.4.** A configuration matrix $X$ has *pre-shape* given by

$$Z = \frac{X_H}{\|X_H\|} = \frac{HX}{\|HX\|}$$

which is invariant under translation and scaling of $X$.

**Definition 1.5.** The *pre-shape space* (denoted $S_m^k$) is the space of all possible pre-shapes, and since $\|Z\| = 1$, it is a unit hypersphere in $\mathbb{R}^{(k-1)m}$.

The term 'pre-shape' indicates that the fully standardized shape of the configuration has yet to be reached, as rotation has not been removed. This final step is performed by identifying rotations of $Z$ in the pre-shape space.

**Definition 1.6.** The *shape* of a configuration matrix $X$ can be represented by the set

$$[X] = \{Z\Gamma | \Gamma \in SO(m)\}.$$

The set of all such shapes is called the *shape space*, and is denoted $\sum_m^k$.

In more topological language, for configuration $X$ with pre-shape $Z$, rotations of $Z$ on the pre-shape sphere form an *orbit* that corresponds one-to-one with the shape represented by $[X]$. Identifying along such orbits gives the shape space

$$\sum\nolimits_m^k = S_m^k / SO(m)$$

i.e. $\sum_m^k$ is the *quotient space* of $S_m^k$ under the action of $SO(m)$.

The shape of a configuration is actually then, an entire set: the equivalence class formed under the group action of similarity transformations. A simplified pre-shape sphere in shown in Figure 1.1.2.

### 1.1.3 Procrustes Distances

It is desirable to define a distance measure between shapes, and some possible choices of metric for the shape space are given below. For two $k \times m$ configuration matrices $X_1$ and $X_2$, with corresponding pre-shapes $Z_1$ and $Z_2$ define:

**Full Procrustes distance:** When minimized over rotations and scale, the closest Euclidean distance between $Z_1$ and $Z_2$ is the *full Procrustes distance*

$$d_F(X_1, X_2) = \inf_{\Gamma \in SO(m), \beta > 0} \|Z_2 - \beta Z_1 \Gamma\|,$$

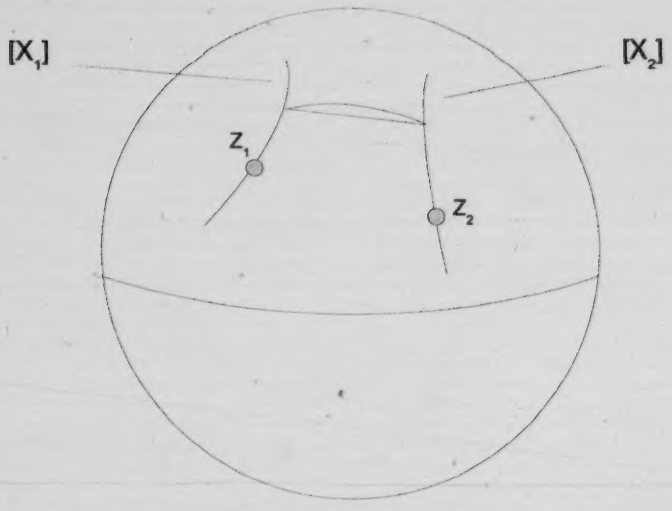


Figure 1.1: A simplified pre-shape sphere showing the orbits corresponding to two shapes.

**Partial Procrustes distance:** The term 'partial' indicates that closest matching is performed over translations and rotations, but not scale. This gives the *partial Procrustes distance* as

$$d_P(X_1, X_2) = \inf_{\Gamma \in SO(m)} \|Z_2 - Z_1\Gamma\|,$$

**Procrustes distance:** Using the pre-shape hypersphere, the great circle distance between $Z_1$ and $Z_2$ minimized over rotations gives the *Procrustes distance* (defined trigonometrically) as

$$\rho(X_1, X_2) = 2\arcsin(d_P(X_1, X_2)/2).$$

Representations of the above distances are shown in Figure 1.1.3, both on the pre-shape sphere and in a cross-sectional view.

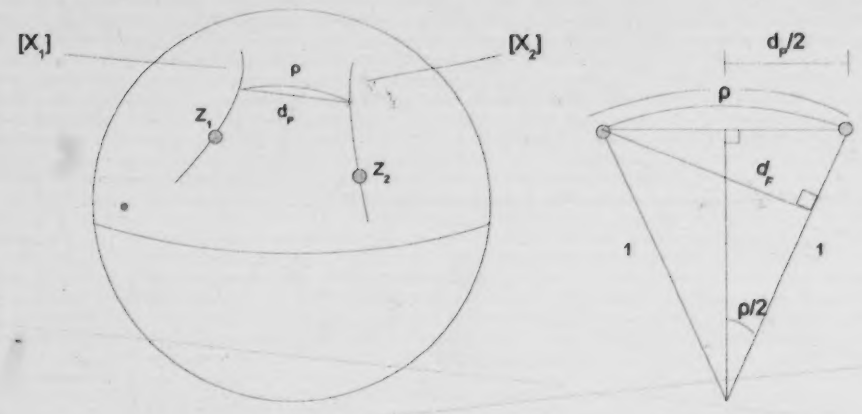


Figure 1.2: On the left is the simplified pre-shape sphere with two of the Procrustes distances added. On the right is a cross-sectional view illustrating the Procrustes distances.

### 1.1.4 Procrustes Methods

We now return to the methods of ordinary and generalized Procrustes analysis (full OPA and full GPA) discussed in Section 1.1.1. Throughout this section, we will assume without loss of generality that all configuration matrices have been centred through pre-multiplication by $C$ (defined in Eq. 1.2).

Full OPA is concerned with matching just one configuration, $X_1$, as closely as possible onto another configuration, $X_2$, by finding the $(\hat{\gamma}, \hat{\beta}, \hat{\Gamma})$ as described in Definition 1.1 that minimize $\|X_2 - \beta X_1\Gamma - \mathbf{1}_k\gamma^T\|^2$.

**Definition 1.7.** When the minimizing transformations are estimated and applied to $X_1$, we denote the transformed configuration with a 'P' superscript, and call it the *full Procrustes fit* of $X_1$ onto $X_2$:

$$X_1^P = \hat{\beta}X_1\hat{\Gamma} + \mathbf{1}_k\hat{\gamma}^T.$$

The parameters $(\hat{\gamma}, \hat{\beta}, \hat{\Gamma})$ are fully computable:

**Theorem 1.1.** *The values of* $(\hat{\gamma}, \hat{\beta}, \hat{\Gamma})$ *that minimize* $\|X_2 - \beta X_1\Gamma - \mathbf{1}_k\gamma^T\|^2$

*are*

$$
\begin{aligned}
(1.3)\ \hat{\gamma} &= 0 \\
\hat{\Gamma} &= UV^T \quad \textit{where} \quad \frac{X_2^T X_1}{\|X_1\|\|X_2\|} = Z_2^T Z_1 = V\Lambda U^T,\ U,V \in SO(m) \\
\hat{\beta} &= \frac{\text{trace}(X_2^T X_1 \hat{\Gamma})}{\text{trace}(X_1^T X_1)},
\end{aligned}
$$

*Proof.* Assuming $X_1$ and $X_2$ centred, the quantity to be minimized is

$$\|X_2 - \beta X_1 \Gamma - \mathbf{1}_k \gamma^T\|^2 = \|X_2\|^2 + \beta^2 \|X_1\|^2 - 2\beta \text{trace}(X_2^T X_1 \Gamma) + k\gamma^T \gamma$$

so $\hat{\gamma}$ is taken to be 0. Now use the singular value decomposition

$$Z_2^T Z_1 = V\Lambda U^T,$$

where $U, V \in SO(m)$ and $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_m)$ with $\lambda_1 \geq \ldots \geq \lambda_{m-1} \geq |\lambda_m|$ the square roots of the eigenvalues of $Z_1^T Z_2 Z_2^T Z_1$, to minimize

$$
\begin{aligned}
\|X_2\|^2 + \beta^2 \|X_1\|^2 - 2\beta \text{trace}(X_2^T X_1 \Gamma) &= \|X_2\|^2 + \beta^2 \|X_1\|^2 \\
(1.4) \qquad & \quad - 2\beta \|X_1\| \|X_2\| \text{trace}(Z_2^T Z_1 \Gamma).
\end{aligned}
$$

Maximizing $\text{trace}(Z_2^T Z_1 \Gamma)$ over $\Gamma \in SO(m)$ gives

$$\hat{\Gamma} = UV^T \tag{1.5}$$

(see Dryden and Mardia [15] p.61-62 for details of Eq. 1.5). Now differentiating the right-hand side of Eq. 1.4 with respect to $\beta$ gives

$$2\beta \|X_1\|^2 - 2\|X_1\| \|X_2\| \text{trace}(Z_2^T Z_1 \hat{\Gamma})$$

so the minimizing value of $\beta$ is

$$\hat{\beta} = \frac{2\text{trace}(\|X_1\|\|X_2\|Z_2^T Z_1 \hat{\Gamma})}{2\|X_1\|^2} = \frac{\text{trace}(X_2^T X_1 \hat{\Gamma})}{\text{trace}(X_1^T X_1)}.$$

as desired. □

In the general case there are $n \geq 2$ configurations that must be superimposed as closely as possible through similarity transformations. For this, generalized Procrustes analysis is used, and the full Procrustes fit will be written using similar notation to the OPA case.

**Definition 1.8.** For each of the configurations $X_i$, the parameters $(\hat{\Gamma}_i, \hat{\beta}_i, \hat{\gamma}_i)$ are those which minimize the sum in Eq. 1.1. Then the fitted configurations

$$X_i^P = \hat{\beta}_i X_i \hat{\Gamma}_i + \mathbf{1}_k \hat{\gamma}_i^T, \ i = 1, \ldots, n,$$

are called the *full Procrustes fit* of the $X_i$.

The model used for the configurations is

$$X_i = \beta_i(\mu + E_i)\Gamma_i + \mathbf{1}_k \gamma_i^T,$$

where the $E_i$ are $k \times m$ independent random error matrices with zero mean, and $\mu$ is the $k \times m$ mean configuration matrix. The similarity transformations are considered *nuisance parameters*, which must be estimated first before estimates for the true parameter of interest $\mu$ can be obtained. As mentioned in Section 1.1.1, full GPA cannot be calculated explicitly for $m > 2$ dimensions, and thus an iterative procedure must be used. The following algorithm, presented in Dryden and Mardia [15], is an adaptation for shape analysis from the full GPA algorithms of Gower [20] and Ten Berge [1].

**GPA Algorithm for $m > 2$**

1. **Translations.** Centre each configuration using centring matrix $C$ given

in Eq 1.2. Initially let

$$X_i^P = X_i \ , \ i = 1, 2, \ldots, n$$

2. **Rotations.** For each $i = 1, \ldots, n$, let

$$\bar{X}_{(i)} = \frac{1}{n-1} \sum_{j \neq i} X_j^P,$$

and the new $X_i^P$ be the rotation-only OPA superimposition of the old $X_i^P$ onto $\bar{X}_{(i)}$. Repeat this until the Procrustes sum of squares

$$(1/n) \sum_{i=1}^{n} \sum_{j=i+1}^{m} \| (\beta_i X_i \Gamma_i + \mathbf{1}_k \gamma_i^T) - (\beta_j X_j \Gamma_j + \mathbf{1}_k \gamma_j^T) \|^2$$

from Eq. 1.1 cannot be reduced further.

3. **Scaling.** The *vectorize* operator, vec$(X)$, takes an $l \times m$ matrix $X$ and stacks the columns of $X$ (in order) to obtain an $lm$-vector. If the vec$(X_i^P)$'s have $n \times n$ correlation matrix $\Phi$ with eigenvector $\phi = (\phi_1, \ldots, \phi_n)^T$ corresponding to the largest eigenvalue of $\Phi$, then take

$$\hat{\beta}_i = \left( \frac{\sum_{k=1}^{n} \| X_k^P \|^2}{\| X_i^P \|^2} \right)^{1/2} \phi_i,$$

for $i = 1, \ldots, n$.

4. **Iteration.** Repeat steps 2 and 3 until the Procrustes sum of squares from Eq. 1.1 can no longer be reduced.

The estimate of mean shape $[\hat{\mu}]$, is the shape with centroid size 1 whose sum of full Procrustes distances to all the $X_i$ is a minimum. In other words:

**Definition 1.9.** The *full Procrustes estimate of mean shape (full Procrustes*

*mean)* is given by $[\hat{\mu}]$, where

$$\hat{\mu} = \arg \inf_{\mu: S(\mu)=1} \sum_{i=1}^{n} d_P^2(X_i, \mu) = \arg \inf_{\mu: S(\mu)=1} \sum_{i=1}^{n} \sin^2 \rho(X_i, \mu).$$

Note that the shape of $\hat{\mu}$ is the same shape as the mean of the full Procrustes fits

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i^P.$$

This is because GPA is performed by minimizing the sum of the squared Euclidean distances, and the minimum of

$$\sum \|X_i^P - \mu\|^2$$

is given by $\mu = \bar{X}$.

For both ordinary and generalized Procrustes analysis, it is possible to perform a version that doesn't allow scaling transformations. This is desirable when size information about the objects of interest is relevant, but the objects are still of comparable scale. This type of Procrustes analysis is called *partial* OPA or *partial* GPA. The sum of squares to be minimized are exactly as those for full OPA and GPA, but with the $\beta$'s removed. The methods of estimation remain the same, with the same translation and rotation estimates in OPA, and Step 3 omitted from the GPA algorithm.

If all the shapes of interest lie reasonably close together in the shape space (i.e. there is not a large amount of variability between the shapes), then the choice of whether to use the Procrustes distance $d_F, d_P$, or $\rho$ may be unnecessary. In the vicinity of the pole (often chosen to be the mean shape), these three distances are very similar, and may be closely approximated through a tangent space projection.

### 1.1.5 Tangent Space Methods

Tangent space approximations are useful, because linearization of the shape space potentially allows for the use of regular multivariate statistics. The method of tangent decomposition discussed here does require some scaling of the original configurations, since the pre-shapes $Z_i$ are used. However, the coordinates are still termed *partial* Procrustes tangent coordinates because the pre-shapes are matched to the pole only through rotation (and not scaling). Doing this allows all the points representing the shapes of interest to remain on the unit hypersphere in $\mathbb{R}^{(k-1)m}$.

The tangent space projection is taken to the pre-shape sphere, using $\hat{\mu}$ as the pole. As with any tangent projection, the tangent vector representing a point in the shape space must be orthogonal to the pole. i.e.) Tangent vector $V$ must satisfy

$$\text{trace}(V^T\mu) = 0,\ \text{trace}(V^T\mu S) = 0. \tag{1.6}$$

for any skew symmetric matrix $S$.

A $(k-1) \times m$ pre-shape $Z$ is rotated to match up with pole $\hat{\mu}$ using $\hat{\Gamma}$ from Eq. 1.3. The vectorize operator $\text{vec}(X)$ defined in step 3 of the GPA algorithm, is used to obtain $v$, the tangent plane projection of $Z\hat{\Gamma}$ at $\hat{\mu}$,

$$v = (I_{(k-1)m} - \text{vec}(\hat{\mu})\text{vec}(\hat{\mu})^T)\text{vec}(Z\hat{\Gamma}). \tag{1.7}$$

Note that since

$$v^T\text{vec}(\hat{\mu}) = \text{trace}(V^T\hat{\mu}) = 0,\ \text{vec}(v)^T\text{vec}(\hat{\mu}S) = \text{trace}(V^T\hat{\mu}S) = 0$$

Eq. 1.6 is satisfied.

The tangent space to the pre-shape sphere at $\hat{\mu}$ (denoted $T_{\hat{\mu}}S_m^k$) is of dimension $(k-1)m - 1$. From a neighbourhood $U$ of $\hat{\mu}$ in the shape space to the pre-shape sphere there is a local section (call it $\sigma$), with $\sigma[\hat{\mu}] = \hat{\mu}$ and

$\sigma[Z] = Z\hat{\Gamma}$ (i.e. it gives the point on the orbit of the pre-shape with minimum distance to $\hat{\mu}$). The image of $U$ under $\sigma$ is a submanifold of $S_m^k$ of dimension $M = (k-1)m - 1 - \frac{m(m-1)}{2}$. Since it is assumed that the shapes are reasonably similar to each other (and to $\hat{\mu}$), then the projection of points in $\sigma(U)$ to $T_{\hat{\mu}}S_m^k$ is, for practical purposes, in $T_{\hat{\mu}}\sigma(U)$. This is a subvector space of $T_{\hat{\mu}}S_m^k$ of dimension $M$, so although the $v_i$'s are written as $(k-1)m$-length vectors, they lie in a space of dimension $M < (k-1)m$.

If scaling of the pre-shapes is desired for fitting to the pole, then full Procrustes tangent coordinates

$$v_F = (I_{(k-1)m} - \text{vec}(\hat{\mu})\text{vec}(\hat{\mu})^T)\text{vec}(\hat{\beta}Z\hat{\Gamma}) \tag{1.8}$$

can be obtained, using $\hat{\beta}$ from Eq. 1.3 as in full OPA.

The tangent space setting allows statistical inferences to be made about the groups of shapes under study. See Dryden and Mardia [15] Chapter 7 for details of Goodall's $F$ test (used under the assumption of isotropy), plus the one-sample analog to the two-sample Hotelling's $T^2$ test discussed below.

*Two independent sample Hotelling's $T^2$ test.*

For two independent random samples $X_1, X_2, \ldots, X_{n_1}$ and $Y_1, Y_2, \ldots, Y_{n_2}$, from populations with mean shapes $[\hat{\mu_1}]$ and $[\hat{\mu_2}]$, test

$$H_0 : [\mu_1] = [\mu_2] \text{ versus } H_1 : [\hat{\mu}_1] \neq [\hat{\mu}_2] \tag{1.9}$$

using the pooled full Procrustes mean shape $\hat{\mu}$ as the pole. With partial Procrustes tangent coordinates of $v_1, \ldots, v_{n_1}$ and $w_1, \ldots, w_{n_2}$ and a multivariate normal model in the tangent space, then for

$$v_i \sim N(\xi_1, \Sigma)\ ,\ w_j \sim N(\xi_2, \Sigma),\ i = 1, \ldots, n_1; j = 1, \ldots, n_2,$$

the usual test statistic for Hotelling's $T^2$ test is the Mahalanobis distance

$$D^2 = (\bar{v} - \bar{w})^T S_u^{-1} (\bar{v} - \bar{w}),$$

where $S_u$ is the pooled sample covariance matrix (with divisor $n_1+n_2-2$). However, because the tangent vectors lie in an $M$-dimensional subspace of $\mathbb{R}^{(k-1)m}$, where $M=(k-1)m-1-\frac{m(m-1)}{2}<(k-1)m$, the covariance matrix $S_u$ will not (in general) be of full rank, making inversion impossible. An adaptation which allows 'inversion' of symmetric square but singular matrices is the Moore-Penrose generalized inverse.

**Definition 1.10.** The *Moore-Penrose generalized inverse* of a square symmetric matrix $A$ is

$$A^{-}=\sum_{j=1}^{p}\lambda_j^{-1}\gamma_j\gamma_j^T,$$

where $\gamma_j$ are the eigenvectors of $A$ corresponding to the $p$ nonzero eigenvalues $\lambda_1,\ldots,\lambda_p$.

Note that $A^-$ satisfies the property that $A^-AA^-=A$.

Using the Moore-Penrose generalized inverse, define the Mahalanobis distance to be

$$D^2=(\bar{v}-\bar{w})^TS_u^-(\bar{v}-\bar{w}).$$

Under $H_0$, $\xi_1=\xi_2$, and using the test statistic

$$F=\frac{n_1n_2(n_1+n_2-M-1)}{(n_1+n_2)(n_1+n_2-2)}D^2$$

(which follows a $F^M_{n_1+n_2-M-1}$ null distribution), $H_0$ is rejected for large values of $F$. If the assumption of equal covariance matrices, or the multivariate normal tangent space model are inappropriate, then a bootstrap procedure may be used. To test the same null hypothesis as in Eq. 1.9, a number $B$ (often between 200 and 1000) is first chosen. Then the total group of $n_1+n_2$ subjects is randomly split into two groups of size $n_1$ and $n_2$ and the test statistic is evaluated for these newly created groups. This is repeated $B$ times, and the ranking $r$ of the observed test statistic $T_{obs}$ compared to those of the $B$ random permutations is used to give a $p$-value:

$$p=1-\frac{r-1}{B+1}.$$

Under the assumption of independent isotropic covariances (i.e. variability is independent and equal along each axis), a one-way analysis of variance (ANOVA) is possible, but the assumption of isotropy is often unrealistic. To test equality of means for several groups a multivariate ANOVA can be performed, using the pooled full Procrustes mean as the pole for tangent space approximations. All usual multivariate analyses for Euclidean space are possible in the tangent space, but a problem arises when the number $n$ of observations is small compared to the dimension of the tangent space. This can occur easily in real-world applications, as medical and biological studies often have a relatively small number of subjects, and $M = (k-1)m - \frac{m(m-1)}{2} - 1$ can become very large as $k$ increases.

One alternative is to apply dimensionality-reducing procedures such as Principal Components Analysis (PCA) or multidimensional scaling (MDS) in the tangent space. Another is the approach we suggest in Chapter 5, to use an approach from computational topology to identify subgroups in the tangent space.

The methods discussed in Section 1.1 will be applied to our 3D orthodontic data set in Chapter 3.

Next we will discuss the method of Euclidean Distance Matrix Analysis, which we will use both as a shape analysis technique, and a method of detecting influential landmarks.

## 1.2 Euclidean Distance Matrix Analysis

This form of shape analysis, usually abbreviated to EDMA, was first presented by Subhash Lele in 1991 [28], [29], and was developed with biological applications in mind. See the book *An Invariant Approach to Statistical Analysis of Shapes* by Lele and Richetsmeier [31] for a full exposition of EDMA and its applications.

### 1.2.1 Description of shape and form

In the EDMA approach, instead of calculating estimates for the nuisance parameters of rotation and translation (as in Dryden and Mardia's approach), each landmark configuration is represented as the set of all pairwise inter-landmark distances. This representation is invariant to these rigid motions. These inter-landmark distances are written in a form matrix. (*Note:* Throughout, the term *form* will be used to refer to size-and-shape information, while the term *shape* refers to information that is invariant to scale)

**Definition 1.11.** For $k \times m$ landmark coordinate matrix $X$, the $k \times k$ *form matrix* of $X$ is

$$FM(X) = \begin{bmatrix} 0 & d_{12} & \cdots & d_{1k} \\ d_{21} & 0 & \cdots & d_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ d_{k1} & d_{k2} & \cdots & 0 \end{bmatrix}$$

where $d_{ij}$ is the Euclidean distance between the $i$th and $j$th landmarks. Note that $FM(X)$ is a square symmetric matrix. Because of the repetition involved in writing $FM(X)$ as a matrix, it is often written as a vector consisting of the above-diagonal entries.

$$FM(X) = [d_{12}, \ldots, d_{1k}, d_{23}, \ldots, d_{2k}, \ldots, d_{(k-1)k}]^T$$

This vector will be of length $l = k(k-1)/2$.

To compare two shapes, each is first represented as a form matrix. One measure of form difference suggested by Lele and Richtsmeier is that of a relative form difference matrix.

**Definition 1.12.** For initial landmark configurations $X$ and $Y$, with form matrices $FM(X)$ and $FM(Y)$, define the *form difference matrix* to be

$$FDM(Y, X) = \left[\frac{FM_{ij}(Y)}{FM_{ij}(X)}\right].$$

i.e. The $(i,j)$th entry of $FDM(Y,X)$ is the ratio of the $(i,j)$th entries of $FM(Y)$ and $FM(X)$ respectively.

As with regular form matrices, an $FDM$ is often written as a vector using the upper-triangle entries.

If the entry of the $FDM$ corresponding to landmarks $i$ and $j$ is greater than 1, then that inter-landmark distance is larger in configuration $Y$ than in configuration $X$ (and smaller if the entry if less than 1).

One measure of the 'size' of an object was presented in Section 1.1: centroid size. In EDMA literature, a common size measure is the geometric mean of all the distances in the form matrix.

**Definition 1.13.** For configuration $X$ with $k \times k$ form matrix $FM(X)$, let the *size* of $X$ be represented by

$$S(X) = \prod FM_{ij}(X)^{1/l}$$

where $l$ is the length of the $FM$ vector (i.e. the number of entries in the upper triangle of the form matrix).

A scaled form of a configuration may now be represented as the form matrix divided by its size, and a *shape difference matrix* defined analogously to the $FDM$.

**Definition 1.14.** For $X$, with $FM(X)$, and $S(X)$ as above, define the *shape matrix* of $X$ to be

$$SM(X) = \left[\frac{FM_{ij}(X)}{S(X)}\right].$$

If configuration $Y$ has shape matrix $SM(Y)$, then the *size difference* between $X$ and $Y$ is the quantity

$$S(Y)/S(X)$$

and the *shape difference matrix* is defined entry-wise as

$$SDM_{ij}(Y,X) = SM_{ij}(Y) - SM_{ij}(X).$$

### 1.2.2 Calculating mean form matrix

In the statistical setting, where a group of objects is taken to be a random sample from some population, we are again interested in estimating the mean shape of the population. Because of the invariance of shape to similarity transformations, the mean shape cannot be estimated directly, so it's form matrix is estimated, and then one instance of that shape (perhaps centred at the origin) is chosen as an *icon* for graphical purposes.

To estimate the population mean shape, begin by writing each landmark configuration $X_1, \ldots, X_n$ as the vector $FM(X_i)$ from Definition 1.11. The method of estimation varies depending on assumptions made about the variance about the mean shape. In general, the covariance matrix $V$ can be written as

$$V = \Sigma_k \otimes \Sigma_m,$$

the Kronecker product of the covariance matrix between landmarks ($\Sigma_k$) and the covariance matrix between the axes ($\Sigma_m$).

**Note:** The *Kronecker product* of an $m \times n$ matrix $A$ and a $p \times q$ matrix $B$, is the $mp \times nq$ matrix

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \vdots & \vdots \\ a_{m1}B & \ldots & a_{mn}B \end{bmatrix}.$$

Under the full model, $\Sigma_k$ and $\Sigma_m$ are symmetric matrices (not equal to the identity matrix) of size $k \times k$ and $m \times m$ respectively. This means the variability at different landmarks is correlated, as is the variability along the $x$, $y$, and $z$ axes. Estimation of the mean form matrix is possible in this case, but can become very computationally intensive as $k$ increases. It involves the solution to a rather complicated system of $k(k+2) + m - 2$ equations. See Lele and Richtsmeier [31] p.117-120 for details.

A slightly less realistic, but more tractable model is $V = \Sigma_k \otimes I_m$, where

the axes are uncorrelated with equal variability. The algorithm for this case from Lele and Richtsmeier [31] is given below.

The model $V = I_k \otimes I_m = I_{km}$ (the isotropic model) is unrealistic in a biological setting, and will not be discussed further here.

**Algorithm to estimate mean form (when $V = \Sigma_k \otimes I_m$):**

1. For $X_1, \ldots, X_n$, calculate the form matrices $FM(X_1), \ldots, FM(X_n)$.

2. Calculate the squared Euclidean distance matrices $E_1, \ldots, E_n$, where the $(l, m)$th entry of $E_i$ is $e_{lm} = [FM_{lm}(X_i)]^2$.

3. For each squared inter-landmark distance (between landmarks $l$ and $m$, say), calculate

$$\overline{e}_{lm} = \frac{1}{n}\sum_{i=1}^{n} e_{lm,i}, \text{ and } s^2(e_{lm}) = \frac{1}{n}\sum_{i=1}^{n}(e_{lm,i} - \overline{e}_{lm})^2.$$

   (The average and variance, over the $n$ individuals, of the squared Euclidean distances between landmarks $l$ and $m$.)

4. Calculate $\hat{\varepsilon}_{lm} = (\overline{e}^2_{lm} - \frac{3}{2}s^2(e_{lm}))^{1/2}$ for all pairs of landmarks $l = 1, \ldots, k; m = 1, \ldots, k$, and use

$$FM(\hat{M}) = \begin{bmatrix} 0 & \sqrt{\hat{\varepsilon}_{12}} & \cdots & \sqrt{\hat{\varepsilon}_{1k}} \\ \sqrt{\hat{\varepsilon}_{12}} & 0 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \sqrt{\hat{\varepsilon}_{1k}} & \sqrt{\hat{\varepsilon}_{1(k-1)}} & \cdots & 0 \end{bmatrix}$$

   as an initial estimate of the mean form matrix. This estimate can be improved by restricting it to be a form matrix for some actual $m$-dimensional object:

5. Square the entries in $FM(\hat{M})$ to obtain

$$EM(\hat{M}) = \begin{bmatrix} 0 & \hat{\varepsilon}_{12} & \cdots & \hat{\varepsilon}_{1k} \\ \hat{\varepsilon}_{12} & 0 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \hat{\varepsilon}_{1k} & \hat{\varepsilon}_{1(k-1)} & \cdots & 0 \end{bmatrix}.$$

Use $EM(\hat{M})$ along with $C = I_k - \frac{1}{k}1_k 1_k^T$ to calculate

$$B(\hat{M}) = -\frac{1}{2}C[EM(\hat{M})]C^T.$$

6. Calculate the eigenvalues $\lambda_1 \geq \ldots \geq \lambda_k$ and corresponding eigenvectors $h_1, \ldots, h_k$ of $B(\hat{M})$.

7. For three-dimensional data, estimate the mean form matrix by

$$\hat{M} = [\sqrt{\lambda_1}h_1, \sqrt{\lambda_2}h_2, \sqrt{\lambda_3}h_3].$$

This configuration can be used if the mean shape needs to be displayed graphically.

8. Use the form matrix corresponding to $\hat{M}$ as the improved estimator for $FM(\hat{M})$.

**Note:** This estimate is accurate only up to rotation, translation *and* reflection, so it is possible that one or more axes may need to be multiplied by $-1$ to obtain an accurate graphical representation.

### 1.2.3 Comparison between groups

To compare two groups of configurations $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$, each configuration is represented as a form matrix, and the two mean form matrices $FM(\hat{X})$ and $FM(\hat{Y})$ are estimated. An estimated mean form difference

matrix is then calculated by substituting the estimated mean form matrices into Definition 1.12.

**Definition 1.15.** The *estimated mean form difference matrix* is given by

$$FDM(\hat{Y}, \hat{X}) = \left[\frac{FM_{ij}(\hat{Y})}{FM_{ij}(\hat{X})}\right].$$

The estimated mean shape matrices and mean shape difference matrix are defined similarly:

$$SM(\hat{X}) = \left[\frac{FM_{ij}(\hat{X})}{S(\hat{X})}\right], \; SDM_{ij}(\hat{Y}, \hat{X}) = SM_{ij}(\hat{Y}) - SM_{ij}(\hat{X}),$$

where $S(\hat{X}) = \left[\prod FM_{ij}(\hat{X})\right]^{1/l}$.

In terms of statistical analysis, it is desirable to obtain point estimates and confidence intervals to discern in which area(s) shape differences are localized. Observing the estimated $FDM(\hat{Y}, \hat{X})$, a point estimate for the linear distance between landmarks $i$ and $j$ does not give very much information. Without knowledge about what type of variability is expected, it cannot be determined whether the given distance is *significantly* different between group $X$ and group $Y$. Although it is possible to obtain confidence intervals using a model-based Monte-Carlo approach (for instance, assuming a Gaussian perturbation model and generating samples based on estimated mean and covariance structure), such methods as sensitive to deviations from the model assumptions. If sample size is not unreasonably small, a bootstrap procedure that doesn't require assumptions about the underlying model is often more appropriate:

**Bootstrap procedure for $FDM$ entry-wise confidence intervals**

The groups of interest are $X_1, \ldots, X_{n_1}$ and $Y_1, \ldots, Y_{n_2}$, with $FDM(\hat{Y}, \hat{X})$ written as a vector of length $l = k(k-1)/2$.

1. From the sample $X_1, \ldots, X_{n_1}$ obtain a simple random sample with replacement, of size $n_1$. Call this sample $X*_1, \ldots, X*_{n_1}$. Obtain a simple

random sample with replacement of size $n_2$ from $Y_1, \ldots, Y_{n_2}$, and call it $Y*_1, \ldots, Y*_{n_2}$.

2. Calculate $FDM(\hat{Y}*, \hat{X}*)$ from the samples obtained in step 1, and write as an $l$-vector.

3. Repeat steps 1 and 2, $B$ times, where $B$ is often chosen between 200 and 1000.

All $B$ of the $FDM(\hat{Y}, \hat{X})$'s can be collected into an $l \times B$ matrix. To obtain a confidence interval for each inter-landmark distance, sort the row containing the $B$ entries for that distance in ascending order. Removing the first $\alpha/2$ and last $\alpha/2$, the remaining maximum and minimum entries give a $100(1-\alpha)\%$ confidence interval for that inter-landmark distance. If the confidence interval contains 1 then there is no significant difference in that linear distance between groups $X$ and $Y$ (at the level $\alpha$). This sorting procedure to obtain confidence intervals is done separately for each row in the $l \times B$ matrix.

In addition to determining the individual inter-landmark distances that appear to be bigger or smaller in one group over the other, a hypothesis test can be performed to determine whether the overall mean shapes are significantly different. The two methods that have been developed for the EDMA setting (and are presented in Lele and Richtsmeier [31]) are the EDMA-I and EDMA-II hypothesis tests.

**EDMA-I hypothesis test** The EDMA-I procedure is a one-way test for equality of mean form, and is performed using samples $X_1, \ldots, X_{n_1}$, and $Y_1, \ldots, Y_{n_2}$.

1. Estimate the mean form matrices for samples $X$ and $Y$, calculate $FDM(\hat{Y}, \hat{X})$, and sort its entries in descending order.

2. Use as a test statistic $T$: the ratio of the maximum entry in the $FDM$ to the minimum entry. If the mean forms of $X$ and $Y$ are very similar,

then all the entries in $FDM(\hat{Y}, \hat{X})$ will be very close to 1 (or some other constant if there is size but no shape difference), and thus the test statistic $T$ will also be close to 1. The further $T$ is from one, the more dissimilar the two mean forms. Using the samples $X$ and $Y$, calculate

$$T_{obs} = \frac{\max_{ij} FDM_{ij}(\hat{Y}, \hat{X})}{\min_{ij} FDM_{ij}(\hat{Y}, \hat{X})}.$$

3. Choose one of the samples as a baseline (say sample $Y$). From this baseline sample, perform a simple random sample with replacement of size $n_1$, and call it sample $X*$. Perform another simple random sample from the same baseline, this time of size $n_2$. Call it sample $Y*$. Perform Steps 1, 2, and 3 using $X*$ and $Y*$ to obtain a new $T$ value.

4. Repeat Step 3 $B$ times, with $B$ between 200 and 1000.

5. The values of $T$ obtained in Steps 3 and 4 can be used to form a histogram of the null distribution of $T$. If $T_{obs}$ lies beyond the upper or lower $\alpha/2$ tails of the distribution, then the null hypothesis may be rejected at level $\alpha$ significance.

**EDMA-II two-way hypothesis test** For samples $X$ and $Y$ as above, the EDMA-II method uses scaling factors on the mean form matrices to perform a two-way test for equality of mean shape.

1. Estimate the mean form matrices for samples $X$ and $Y$, and calculate $FDM(\hat{Y}, \hat{X})$, as well as the sizes $S(\hat{X})$ and $S(\hat{Y})$ from Definition 1.13.

2. Obtain the estimated shape difference matrix $SDM(\hat{Y}, \hat{X})$ from Definition 1.14, using the shape matrices $SM(\hat{X})$ and $SM(\hat{Y})$ (where a shape matrix is the $FDM$ divided by size).

3. Using estimates for the mean forms and variance-covariance matrices (see Lele and Richtsmeier [31] p116-120 for variance-covariance matrix estimation) generate two new samples under the Gaussian perturbation model. Calculate the $Z$ statistic for these samples.

4. Repeat Step 3 $B$ times (eg. $B = 500$) to obtain $B$ $Z$-statistics $Z_1, \ldots, Z_{500}$. Sort these in increasing order, remove the top and bottom $\alpha/2$ values, and use the remaining maximum and minimum $Z$ values as the bounds for a $100(1-\alpha)\%$ confidence interval. If this interval does not contain 0, then the null hypothesis may be rejected, and the conclusion made that the samples have significantly different mean shape.

Each test has its advantages and disadvantages. EDMA-I requires the choice one of the samples as a baseline, so it is a one-way test, asking whether the mean shape of group $X$ is similar to a multiple of the mean shape of group $Y$. i.e.) Could sample $X$ have arisen from the same distribution from which sample $Y$ was obtained? Since it is not a symmetric test, it is useful to perform it twice, using the other sample as baseline on the second test. Additionally, EDMA-I assumes equal variance for the two populations, but does not make any assumptions about the underlying distributions.

EDMA-II does not assume equal variances, and is a two-way test that doesn't require the choice of a baseline sample. It does however use a Monte-Carlo method for generating samples to obtain confidence intervals, which assumes a Gaussian perturbation model and requires estimation of variance-covariance matrices. It also requires a choice of size measure (usually the geometric mean from Definition 1.13) to use when standardizing for scale.

Some of the EDMA-based shape analysis methods discussed in the section will be applied to our example data set in Chapter 3.

### 1.2.4 Using EDMA to identify influential landmarks

In the previous section a bootstrap procedure on the $FDM$ was discussed, as a method of determining which inter-landmark distances were significantly different between two mean forms. Other methods involving the $FDM$ are also available, which allow detection of the landmarks that reflect the greatest shape differences between the samples. Such landmarks are termed *influential*.

These methods were first presented by Lele and Richtsmeier in [30].

One method of detecting influential landmarks is to regard the vectorized estimated form difference matrix obtained from the original samples of $k \times m$ configurations $X_1, \ldots X_{n_1}$, and $Y_1, \ldots, Y_{n_2}$. The entries of the estimated $FDM$ vector may be sorted from largest to smallest, and checked to see if any individual landmarks tend to have a lot of associated $FDM$ entries that are far from 1. Similarly, if any region that is biologically/clinically/functionally relevant tends to be involved with the extreme $FDM$ entries, this could indicate that the given region contributes more to the shape difference.

Another method is one that requires a bit more computation, but gives a more quantitative measure of how influential given landmarks are. This method is the *landmark deletion approach*. It uses the $T$ statistic defined in the EDMA-I description,

$$T = \frac{max_{ij} FDM_{ij}(\hat{Y}, \hat{X})}{min_{ij} FDM_{ij}(\hat{Y}, \hat{X})},$$

which is close to 1 when the estimated shapes are similar, and further from 1 the more dissimilar they are. The landmark deletion method consists of calculating the test statistic for the full configurations, and then for $i = 1, \ldots, k$ deleting the $i$th landmark and re-calculating the value of $T$ (which will be denoted $T_{(-i)}$). After obtaining $T_{(-1)}, \ldots, T_{(-k)}$, compare each to the original $T$. The landmark deletions that create the largest reduction in the $T$ value can be considered the most influential. In other words, if the estimated shapes appear to be much more similar (as measured by the $T$ statistic) after the removal of landmark $i$, then the area around landmark $i$ likely accounts for a large proportion of the shape difference. If there are areas, or clusters of landmarks, that are deemed by researchers to be of interest, then they can be deleted in a similar manner, and the $T$ statistic re-calculated. The relative decrease in the $T$ statistic indicates how influential the cluster of landmarks is. This method of detecting influential landmarks will be used in Chapter 4 on the example data set.

## 1.3 Other methods in shape analysis

A number of other methods in shape analysis have been developed, which are beyond the scope of this text. This section will give a partial list of these methods, and references for further reading.

The method of *thin-plate spline analysis* is similar to the deformation grids proposed by Thompson in 1917 [43], and was developed mathematically by Bookstein [5], [6]. It involves fitting one shape (or mean shape), as represented by a landmark configuration, onto another through a combination of affine and non-affine transformations, while minimizing the bending energy required to perform the non-affine transformation. The deformation between configurations may be represented visually as a deformed grid. When the objects of interest are two-dimensional these grids are easily displayed, but for three dimensions the warped grid representation may not be as clearly viewable. In addition to the Bookstein references above, thin-plate spline methods are discussed in [41], [7], and [15].

Another method used to compare two landmark-based shapes (or mean shapes) is *finite element analysis*. During this approach, which was developed from similar methods used in engineering applications, the landmarks are used to discretize the shape into a number of smaller components (usually triangles or quadrilaterals in two dimensions, and tetrahedra or cuboids in three dimensions). The strain involved in transforming one configuration into another is used to describe the shape difference between the objects. For each discretized component (called a finite element) the forces acting at the 'nodes' (i.e. landmarks) are used to infer the forces throughout the rest of the component. Finite element methods provide readily interpretable graphical output (with each element colour-coded according to its level of size/shape change), but do not allow the comparison of multiple forms simultaneously. Presentations of general finite element methods are available in [47] and [45], with discussions specific to shape analysis in [13] and [36].

For some types of shapes, landmark data is impractical. For example, an

organ of the body might not have any well-defined points or edges (and is instead more of an elongated blob). One method that is particularly useful for such cases is *medial axis analysis*. To undergo medial axis analysis a shape must be represented by its boundary. This boundary is then represented by a medial axis, which consists of points which are midway between two points on the boundary (i.e. a sphere centered at a point on the medial axis will be bitangent to the boundary at two points). If full re-constructability of the object from the medial representation is desired, then the radius of the bitangent sphere at each radial axis point is also retained. At points which are equidistant to more than two boundary points, the medial axis branches. For two-dimensional shapes the medial axis is a branching line, and for three dimensions is it a sheet. The method was first presented by Blum [2], and a book on the topic by Pizer and Siddiqi is scheduled for publication in fall 2008 [39].

Other boundary-based methods are those that use Fourier descriptors to make successive approximations of the boundary, using digitized points along the boundary (which can be chosen as specific landmarks, but this is not required). Harmonic coefficients are calculated in a stepwise fashion, with the number of harmonics constrained (under the Nyquist frequency) to be half the number of available boundary points: the more boundary points, the closer the approximation. Initially presented in [46] and [44], see [32] for an example of applications.

Another group of researchers has been using an approach based on *deformable templates*, which model objects under the action of some group (which in shape analysis might be translations/rotations/scaling). Diffeomorphisms are used to match objects, either through landmark matching using large deformation diffeomorphisms, surface matching using various methods (including currents), and matching of entire voxelated 3D images through geodesics on infinite-dimensional shape manifolds. See for example [35], [25] and [34] for further details, and also the book *Pattern Theory: From Representation to Inference* by Grenander and Miller [21].

Finally, the use of persistent homology methods (that will be applied in Chapter 5 to the tangent space data) have been employed as shape descriptors for objects represented as point cloud data. This method is described in [14], and [12]. Another method involving algebraic topological tools as shape descriptors is given in [10].

# Chapter 2

# Landmark reliability

This chapter begins with a description of the data set which will be utilized throughout the remainder of the text. Following that is a discussion of the reliability of landmark placement, and its implications on further analysis, as well as methods to determine landmark reliability in the shape analysis setting.

## 2.1 Description of data set

Throughout the rest of this text a data set from the field of orthodontics will be used to illustrate various statistical shape analysis techniques. A study was undertaken to determine the effects of a bone-anchored orthodontic treatment, as compared to a tooth-anchored treatment and a control group (no treatment). The study was performed at the University of Alberta's Orthodontic Clinic, by Manuel Lagravere (ML) and Paul Major (PM). A total of $n = 62$ subjects participated in the study, and were assigned randomly into one of the three treatment groups ($B$ = bone-anchored, $T$ = tooth-anchored, and $C$ = control) with $n_B = 21$, $n_C = 21$, and $n_T = 20$.

The subjects were followed throughout the course of treatment, with each subject measured at four time points:

1. Prior to placement of orthodontic appliance (baseline).
2. Midway through treatment.

3. At removal of appliance.

4. Six months after removal.

All four time points are of interest, but for purposes which require a before/after comparison, time points T1 and T3 will be used. In total there are 248 landmark configurations (4 for each of the 62 subjects).

Three-dimensional imaging of the subjects was performed through a cone beam computed tomography (CBCT) scan. As discussed in the previous chapter, the statistical shape analysis tools we will be using take landmark coordinate data as their input. $k = 44$ landmarks were chosen to represent the craniofacial complex (the shape of interest). A description of the landmarks is available in Table A.1. The placement of the landmarks was performed by the same individual (ML) for all the subjects at all time points, and was done on the digitized CBCT images using a software program. It is important to determine whether the placement of the landmarks can be done in a reliable way, because otherwise the results of any further analysis of the landmark configurations can be called into question. Determining the reliability of landmark placement is the subject of the remainder of this chapter.

## 2.2 Intraclass Correlation Coefficients (ICCs) as a measure of reliability

When the term 'reliability' is used here, it is in the statistical sense, so a measurement system that produces nearly identical results when repeated under identical conditions is considered 'reliable' (i.e. there is little random error in the measurement).

To quantify the notion of reliability, when the measurement of a single variable is of interest, the intraclass correlation coefficient (ICC) is often used. Depending on the statistical model used in the reliability study, different types of error can be taken into account, and different formulas to calculate the ICC are used correspondingly (as presented in Shrout and Fleiss [42], and further

in McGraw and Wong [33]). All variations attempt to estimate the proportion of total variance that is due to 'true' variation in the objects/measurements of interest. This can generally be stated as

$$ICC = \frac{\text{subject variance}}{\text{subject variance} + \text{ rater variance } + \text{error}} \tag{2.1}$$

A reliability study was performed using just 24 of the landmark configurations (12 chosen randomly from time point 1 and another 12 chosen randomly from time point 2). For this study, landmark placement was replicated on the same set of CBCT scans: three times with ML as rater (ML1, ML2, and ML3), and twice more by two other raters: PM and Jill Gordon (JG). The final study will only use ML as a rater, so his reliability is of primary importance, but inter-examiner reliability between ML, PM and JG is of interest for future studies, so was included here. Since the landmark placement was performed on the same images (and not on multiple CBCT scans of the same subject at the same time point), any changes in coordinates is due to variation in landmark placement, and not the image itself. The results do not need to be able to be extended to any population of raters, so the form of ICC that corresponds to a two-way mixed effects model can be used (fixed raters, but subjects chosen randomly from the subjects to be used in the final study):

$$ICC = \frac{BMS - EMS}{BMS + (j-1)EMS} \tag{2.2}$$

where BMS is the mean square between subjects, EMS is the residual mean square (which can both be obtained from a two-way mixed effects model ANOVA table), and $j$ is the number of judges or raters.

The simplest approach is to calculate the ICC for each of the x- y- and z-coordinates for each landmark. This gives a total of $k \times m = 44 \times 3 = 132$ intraclass correlation coefficients. They are calculated once using ML's three replications (see Table A.2), and again using ML1, PM and JG (see Table A.3). Summary statistics and histograms of the ICCs are displayed in Figure 2.2.

| | ICC summary statistics | | | | | |
|---|---|---|---|---|---|---|
| | Min. | 1st Qu. | Median | Mean | 3rd Qu. | Max. |
| Intra-examiner | 0.9623 | 0.9953 | 0.9970 | 0.9948 | 0.9982 | 0.9998 |
| Inter-examiner | 0.7656 | 0.9875 | 0.9929 | 0.9844 | 0.9961 | 0.9995 |

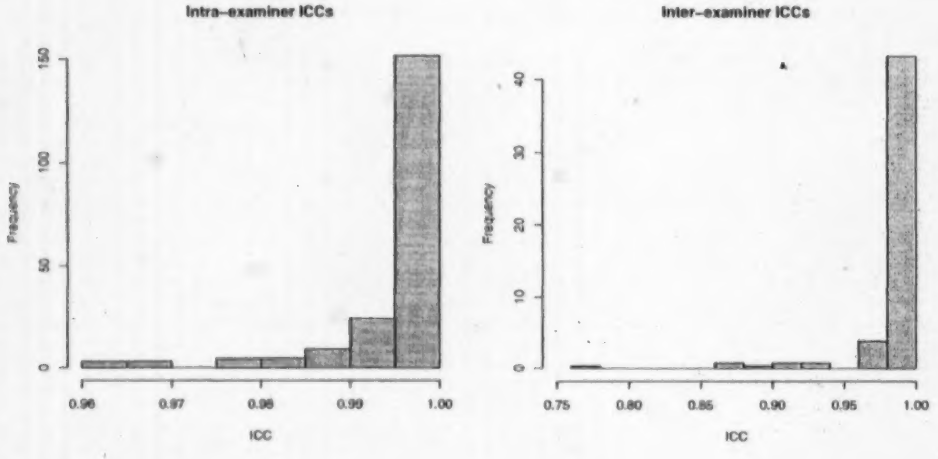


Figure 2.1: Summary statistics and histograms for the 132 ICCs obtained using ML's three replications (intra-examiner), and the 132 using MG1, JG and PM's measurements (inter-examiner).

| | Intra-examiner | | | Inter-examiner | | |
|---|---|---|---|---|---|---|
| | $x$ | $y$ | $z$ | $x$ | $y$ | $z$ |
| ICC | 0.9994871 | 0.9997058 | 0.9988117 | 0.9977847 | 0.999132 | 0.997402 |

Table 2.1: ICCs using all subjects and all landmarks. Given separately for $x$-, $y$-, and $z$-dimensions.

In general, landmark placement by the single rater appears to be quite a bit more reliable then landmark placement across multiple raters. The intra-examiner ICC is very high for all the axes of all the landmarks, with only 4 coordinates having an ICC below 0.975. For the inter-examiner ICCs, there appears to be one coordinate (the $x$-axis of the 4th landmark) whose reliability is much lower than the others (only 0.766). Upon inspection, there is one configuration which has the placement of this coordinate off by 2cm from one rater to the others. Other configurations also have discrepancies in this coordinate up to 0.7 or 0.8 cm between raters. Since changes of only 0.3 or 0.4 cm can be of clinical significance, this landmark may not be reliable enough for use in future studies with multiple raters. Even when using only one rater, this coordinate also had one of the lowest levels of reliability (ICC=0.966).

Ideally, the reliability of landmark placement for an entire configuration could be measured in a more concise way, instead of by $k \times m$ separate ICC values. One option is to calculate ICCs for each dimension using all landmarks of all the subjects (one ICC for each of the $x$-, $y$- and $z$-dimensions). The problem with this is that the formula for ICC is for univariate data, so it will take each coordinate as one 'subject'. Because the variability between landmarks is much greater than the variability in placement of a single landmark, this will cause the 'subject' variance of Eq. 2.1 to become very large compared to the rater variability. This results in very high ICC values, even if the reliability of individual landmark placement is not good (in terms of clinical significance). The intra- and inter-examiner ICCs using this method are given in Table 2.2

Even calculating ICCs for the entire configuration of just one subject at a time, the large variability between landmarks (compared to small variability

in landmark placement) give very high ICCs.

An alternative attempt to take the entire configuration into account when calculating reliability is the use of Principal Components Analysis (PCA) as a dimensionality-reducing tool. For a data set with a large number of variables, PCA determines the linear combination of those variables that accounts for the *most* variability in the data; this linear combination is the *first principal component*. The *second principal component* is the linear combination of the variables the explains the greatest proportion of the variance, subject to being orthogonal to the first principal component. The third must be orthogonal to the first two, and so on. For details on PCA, refer to any textbook on multivariate statistics, such as Johnson and Wichern [24].

To apply PCA here, the coordinates for each landmark are taken as separate 'variables', giving 44×3=132 in total, with 24 observations on each. The first principal component is obtained for each of the rating occasions. Reliability analysis can now be performed using the PC coefficients as measurements. This method displays the same problem as above: very large ICC values are found, because the values of the principal components (PCs) are due mostly to variability between different landmarks, so small changes in landmark placement do not effect the PCs greatly.

Any method that will be able to detect clinically significant differences in landmark placement (i.e. any differences of more than 0.3cm), cannot use all the landmarks on their original scales (which have differences of 10s of cm). The first method presented, that of calculating separate ICCs for each dimension of each landmark, seems to be a good one.

## 2.3 Dry skull data

To assess the reliability of landmark placement on CBCT images as a way of representing a real-world 3D form, a study involving a set of dry skulls was performed. This data set is also used to illustrate some issues that may arise when calculating reliability based on transformed data. The study used CBCT

images of 10 dry skulls, and 8 foramina (holes) were chosen as the landmarks to represent each. The foramina chosen are described in Table A.4. Foramina were used as landmarks, because they may be filled with gutta-percha (a type of latex used in orthodontics) and re-scanned for use as a 'gold standard' of actual landmark locations.

The 10 skulls were scanned once without gutta-percha, and once with. Landmark placement was performed on the non-gutta-percha images three times by ML (call the measurements ML1, ML2 and ML3), and once each by JG and Carlos Flores-Mir (CF). The raw measurements were compared for intra- and inter-rater reliability (displayed in Table 2.2), and also compared with the landmark placement as determined by the gold standard.

It appears that both intra-examiner and inter-examiner reliability is lowest for landmarks 7 and 8, and highest for landmark 1. Interestingly, the landmark 6 x-coordinate is one of the highest for intra-rater reliability, but near the lowest for inter-rater. Upon inspection, the inter-rater measurements for this coordinate are generally very close, except for in one case, when one rater was 0.35cm away from the other two. When compared to the gold standard, each of the examiners had an average measurement error between 0.5-0.8 for each landmark dimension (see Table A.5).

A transformed version of the dry skull configurations was also available. Each configuration was centred and scaled so that a specified point lay at the origin, with the triangle made by that point and two other specified points constrained to lie in the x-y-plane, and a fourth point in the y-z plane. When comparing ratings to the gold standard, using the transformed data improves the match. All figures are shown using the eight original landmarks, plus the four used for the transformation. Figure 2.3 shows the mean shape on the first rating occasion (ML1) and the mean shape using the gold standard (in each of the three planes) using the raw data. The same in shown in Figure 2.4 using the transformed data. It can be seen in Figure 2.3 that the image registration seems to be a bit shifted between the original scans, and those with the gutta-percha filled foramina. This translation between the original and gold standard

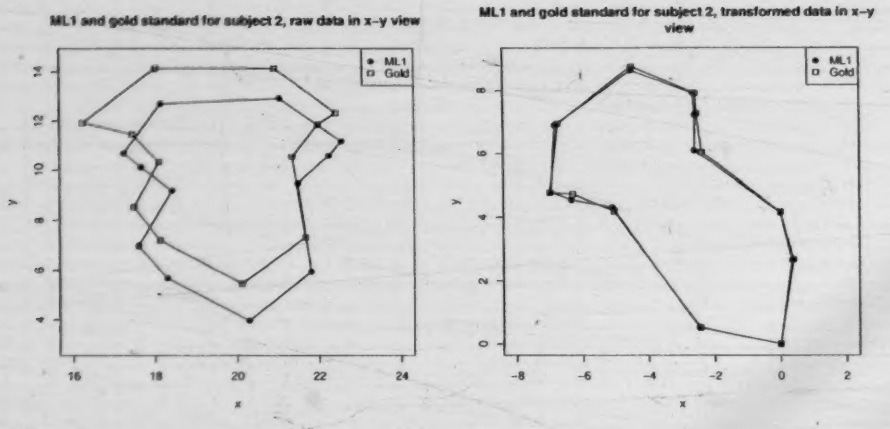


Figure 2.2: Plots comparing the ML1 data to the gold standard for subject 2 using the raw data (left) and transformed data (right).

images is easily seen when looking at individual configurations. Figure 2.2 shows the x-y plane of subject 2 using the raw data for ML1 and the gold standard, as compared to the transformed data. This resulted in the apparent large measurement errors between raters and the gold standard when using the raw data. After transformation, the average measurement error between the examiners and the gold standard is less than 0.2cm (and usually below 0.1cm) for each landmark dimension. Since it is known that ICCs are not invariant to changes in scale, they are obtained again for the transformed skull data (shown in Table 2.3).

When using the transformed data, the reliability of landmark coordinates is related to the axis for both intra- and inter-rater situations. In both cases coordinates on the z-axis are most reliable, and coordinates on the x-axis are least reliable. This is because in the transformed data there is much more variability between the subjects in the z-direction than in the x- and y-directions. In particular, there is one subject whose transformed landmarks lie considerably further in the z-direction than the other subjects. See Figure 2.6: when viewed from the x-y plane it appears that the transformed data has just been translated sideways and rotated a bit, but viewing in the x-z plane it can be seen that the transformed data are much more scattered in the z-

direction from subject to subject. The increased inter-subject variability makes the inter-rater variability smaller in comparison to it. The type of centering and scaling transformation performed on the data should not effect the shape analysis techniques of the following chapters (because shape information is invariant to translations and rotations), but it does effect ICCs, which are dependent on scale. Thus, when measuring reliability using ICCs, the scale and location of the data must be kept in mind.

One final way to check reliability is by calculating ICCs on the Procrustes superimposed landmarks. Procrustes superimposition is performed using all twelve landmarks (holding none fixed). The ICCs are shown in Table 2.3, and give lower reliability scores than the raw or transformed data on the eight original landmarks. This is likely due to a decrease in the between-subject variance after superimposition. See Figure 2.7 for an x-y plane view of the landmarks before and after superimposition. Clearly each landmark displays much more variability between subjects before superimposition. The four 'fixed' landmarks show very high ICCs even after Procrustes superimposition. This lends support to the assumption that these landmarks can be placed very reliably. Plots comparing the mean shape of ML1 to the gold standard after Procrustes superimposition are given in Figure 2.5. Comparing ML1 to the gold standard for the Procrustes superimposed data gives similar measurement error as the transformed data (around 0.1cm).

Each of the above settings for calculating ICCs have somewhat different results. It is important to be aware that ICCs are a measure of variability between ratings *as compared to variability between subjects*, and thus are sensitive to any transformation or superimposition methods that are meant to standardize the data.

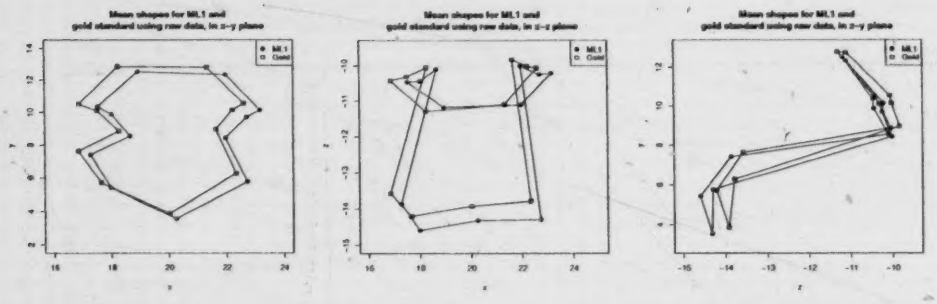

Figure 2.3: Plots comparing the mean shape from the first rating occasion (ML1) to the gold standard, using the raw data. Views from the x-y (left), x-z (middle), and z-y (right) planes.

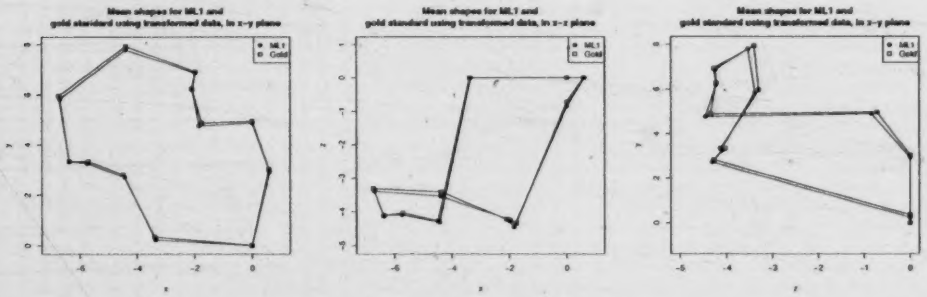

Figure 2.4: Plots comparing the mean shape from the first rating occasion (ML1) to the gold standard, using the transformed data. Views from the x-y (left), x-z (middle), and z-y (right) planes.

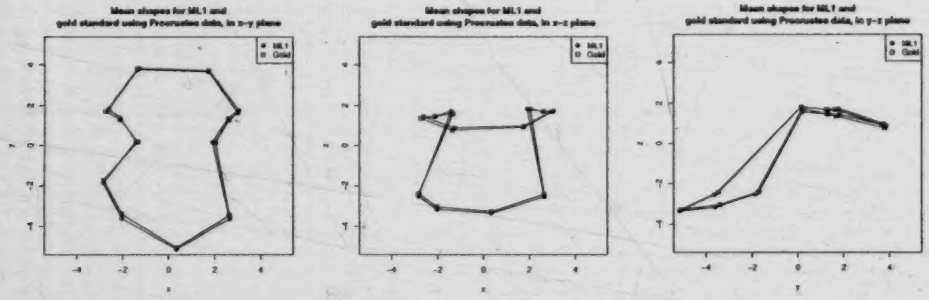

Figure 2.5: Plots comparing the mean shape from the first rating occasion (ML1) to the gold standard, using the Procrustes superimposed data. Views from the x-y (left), x-z (middle), and z-y (right) planes.

| | Intra-rater ICCs | | | Inter-rater ICCs | | |
|---|---|---|---|---|---|---|
| | x | y | z | x | y | z |
| Landmark 1 | 0.994279 | 0.9956914 | 0.9866068 | 0.9872621 | 0.99537 | 0.9218591 |
| Landmark 2 | 0.9489195 | 0.983702 | 0.969373 | 0.9820265 | 0.9861788 | 0.9211727 |
| Landmark 3 | 0.9808405 | 0.9960356 | 0.9920794 | 0.9875333 | 0.9972863 | 0.983666 |
| Landmark 4 | 0.9481057 | 0.9847869 | 0.97811 | 0.9475444 | 0.9670927 | 0.9586 |
| Landmark 5 | 0.994378 | 0.9870798 | 0.9937567 | 0.9962427 | 0.9610836 | 0.9830809 |
| Landmark 6 | 0.996312 | 0.9793826 | 0.9928712 | 0.9460225 | 0.9376382 | 0.9839927 |
| Landmark 7 | 0.9477915 | 0.964017 | 0.9967333 | 0.8223573 | 0.9818152 | 0.930864 |
| Landmark 8 | 0.8016406 | 0.9381783 | 0.9682448 | 0.8292102 | 0.9259185 | 0.9781929 |

Table 2.2: Intra- and inter-rater ICCs for dry skull data.

| | Intra-rater ICCs | | | Inter-rater ICCs | | |
|---|---|---|---|---|---|---|
| | x | y | z | x | y | z |
| Landmark 1 | 0.9979582 | 0.9995924 | 0.9998549 | 0.9984994 | 0.9991692 | 0.9991564 |
| Landmark 2 | 0.9982846 | 0.9991333 | 0.9998512 | 0.9977046 | 0.9996258 | 0.9997027 |
| Landmark 3 | 0.9979015 | 0.999471 | 0.9998887 | 0.9979797 | 0.9997597 | 0.999806 |
| Landmark 4 | 0.9951827 | 0.999564 | 0.9999102 | 0.9949894 | 0.9986471 | 0.9998337 |
| Landmark 5 | 0.995674 | 0.9991747 | 0.9999245 | 0.9907488 | 0.9977368 | 0.9997404 |
| Landmark 6 | 0.9956116 | 0.9994616 | 0.999911 | 0.975422 | 0.9976196 | 0.9997778 |
| Landmark 7 | 0.9945224 | 0.9939348 | 0.9999465 | 0.987474 | 0.9951045 | 0.999073 |
| Landmark 8 | 0.9638145 | 0.9980178 | 0.9997983 | 0.9690613 | 0.9975283 | 0.9998528 |

Table 2.3: Intra- and inter-rater ICCs for transformed dry skull data.

| | Intra-rater | | | Inter-rater | | |
|---|---|---|---|---|---|---|
| | x | y | z | x | y | z |
| Ref point 1 | 0.999577 | 0.9976445 | 0.999795 | 0.9989848 | 0.9971586 | 0.9997074 |
| Ref point 2 | 0.9999472 | 0.9993834 | 0.9981969 | 0.9998255 | 0.9994047 | 0.9987236 |
| Ref point 3 | 0.9999554 | 0.9989747 | 0.9999423 | 0.999881 | 0.998757 | 0.9997373 |
| Ref point 4 | 0.9999363 | 0.9996158 | 0.9996869 | 0.9998024 | 0.9995686 | 0.9996187 |
| Landmark 1 | 0.959817 | 0.965082 | 0.9757493 | 0.9220805 | 0.9511255 | 0.8354375 |
| Landmark 2 | 0.9590607 | 0.9698615 | 0.9918958 | 0.940262 | 0.9554938 | 0.9769037 |
| Landmark 3 | 0.8334562 | 0.9743218 | 0.994149 | 0.8593874 | 0.950443 | 0.990105 |
| Landmark 4 | 0.9458252 | 0.9825304 | 0.9965098 | 0.8996419 | 0.96949 | 0.9924263 |
| Landmark 5 | 0.9119282 | 0.9936243 | 0.9983384 | 0.8972266 | 0.9774856 | 0.9963439 |
| Landmark 6 | 0.9675828 | 0.993492 | 0.9986115 | 0.7065889 | 0.971002 | 0.9968336 |
| Landmark 7 | 0.9089707 | 0.9642265 | 0.999177 | 0.8233037 | 0.9804521 | 0.9909334 |
| Landmark 8 | 0.822423 | 0.9779338 | 0.9973971 | 0.849618 | 0.9732542 | 0.997695 |

Table 2.4: Intra- and inter-rater ICCs for Procrustes superimposed data.

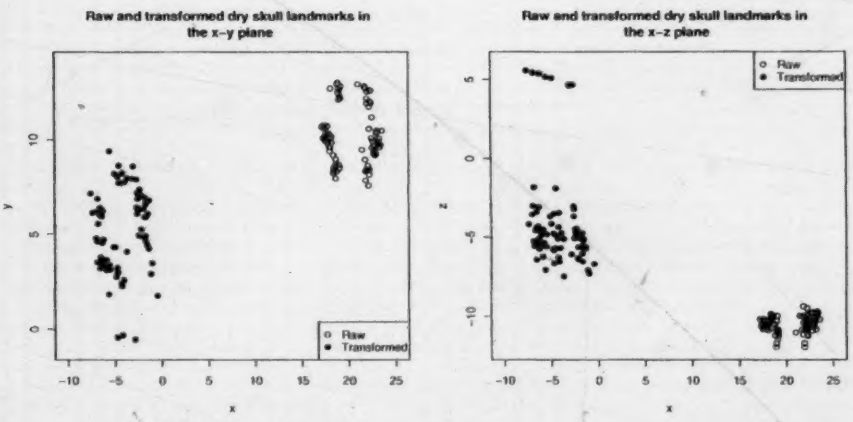


Figure 2.6: Plot of the raw (open circles) and transformed (filled circles) landmarks on the first rating occasion (ML1), viewed from the x-y (left) and x-z (right) planes. A similar relationship is seen on other ratings as well.

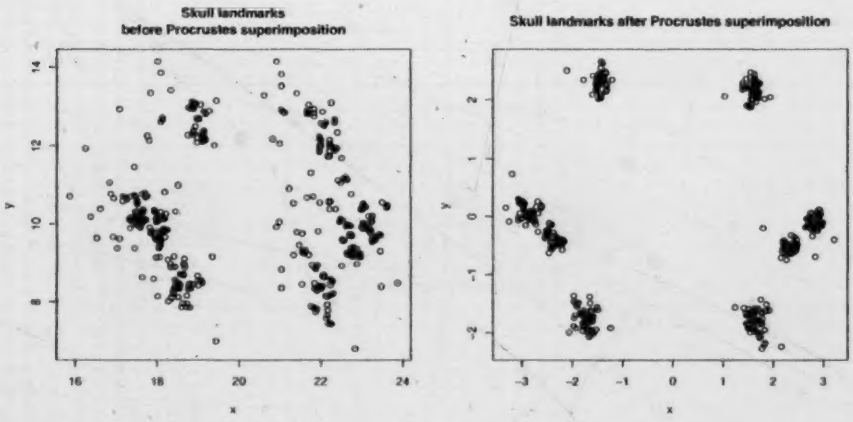


Figure 2.7: Plots of the x-y plane for landmarks before (left) and after (right) Procrustes superimposition.

# Chapter 3

# Application of Shape Analysis Techniques

Now that reliability of landmark placement has been determined, the shape analysis techniques presented in Chapter 1 will be applied to the data set. Initially, the methods will be applied to compare the three treatment groups at baseline, to check for any pre-treatment differences between the groups. Next, each treatment group will be examined over the course of treatment for changes in shape, and the groups will be compared with each other at each time point.

Since the data set consists of three different treatment groups (B, C, and T), at four different time points (1, 2, 3, and 4, as described in Section 2.1), these combinations will be denoted B1, B2, B3, B4, C1, C2, C3, C4, T1, T2, T3, and T4. Each landmark configuration is represented as a $k \times m = 44 \times 3$ configuration matrix. For each group-time point combination the configuration matrices are combined to form an array. For example, the array B1 is a $k \times m \times n_B$ array consisting of the $n_B$ landmark configurations of the subjects in group B at time 1. The arrays B2, ..., T4 are defined similarly. Measurements for the control group were not obtained at time point 2 (since it was assumed no change would occur in the control group), so the configurations for C2 are taken to be identical to those at C1, for computational purposes. Also, landmarks

11 and 12 represent the mesial surface of the left and right incisors (the middle of the top front two teeth), so in patients with no gap these landmarks are identical. From both a shape analysis and clinical perspective it is of interest whether these two landmarks coincide or not, but during EDMA computations this can prove problematic. During EDMA-based portions of this thesis these landmarks will be omitted, and the remaining $k = 42$ will be used. When using other shape analysis methods, the entire 44-landmark configuration will be used.

All of the implementations in this Chapter are performed using the statistical computing package, R. An R-package called *shapes*, containing routines for many of the methods described in Section 1.1 has been written by Ian L Dryden, and was employed here for many of the computations. The R code used in this thesis for EDMA was written by Jennifer Gamble. See Appendix B for some code and functions written for this thesis.

## 3.1 Comparing groups at baseline

This section deals with the groups B1, C1, and T1. Since the measurements for these groups were taken prior to treatment onset, there should be no significant differences in shape between the groups. This assumption will be verified using the shape analysis methods from Chapter 1.

### 3.1.1 Dryden and Mardia methods

Using all 62 of the subjects at baseline, standardization is performed using the generalized Procrustes analysis (GPA) algorithm described on p.11, and an overall mean shape $\hat{\mu}$ is calculated. See Figure 3.1 for the landmarks before and after superimposition. Since none of the subjects in the study exhibit large deformations of the face or skull, all the configurations have a reasonably similar shape, and tangent space approximations to the shape space may be used (with $\hat{\mu}$ as the pole). The method used is the one without scaling of the

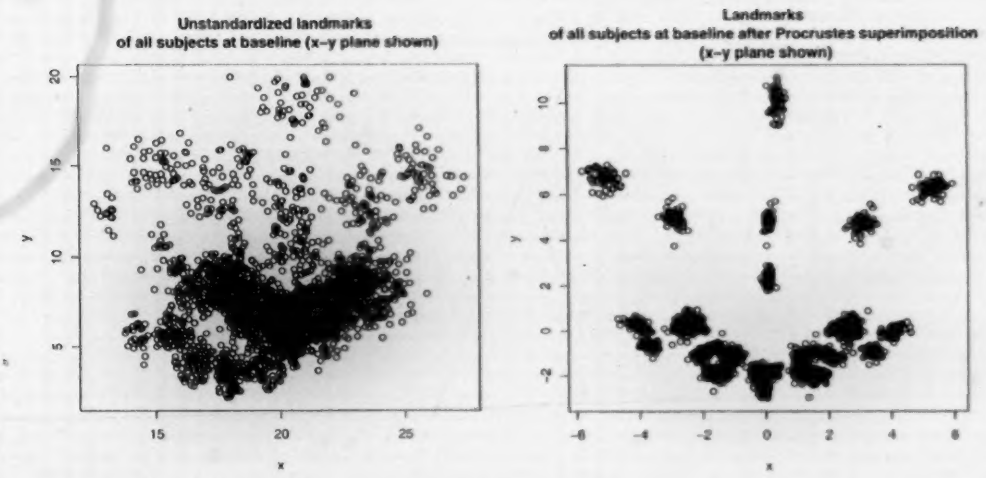


Figure 3.1: x-y view of the landmarks for all 62 subjects at baseline, before GPA was applied (left) and after superimposition (right).

pre-shapes (i.e. Eq. 1.7 instead of Eq. 1.8). After tangent space projection, each configuration is represented as a $(k-1)m = 129$-length vector in the tangent space, which is an $M = (k-1)m - \frac{m(m-1)}{2} - 1 = 125$-dimensional subspace of $\mathbb{R}^{(k-1)m} = \mathbb{R}^{129}$.

The Euclidean distances from each tangent coordinate to the mean shape (i.e. the lengths of the tangent vectors) can be used as a method of detecting outlying shapes. It can be very time consuming to perform data cleaning methods (plotting for outliers, checking descriptive statistics, etc) on the raw landmark configuration data. If a landmark placement or recording error occurs that significantly changes the shape of the configuration, then it will be much further from the mean shape than the other, properly recorded, configurations will be. In fact, for the data set used here, plotting the lengths of the vectors and individually examining any extreme outliers allowed the detection of a configuration with mislabeled landmarks. For that particular subject, during landmark placement the landmarks 39, 40, 41, and 42 had accidentally been labeled as 41, 42, 39, 40. This changed the overall shape of the configuration rather drastically, and it was seen as an outlier in the plot.

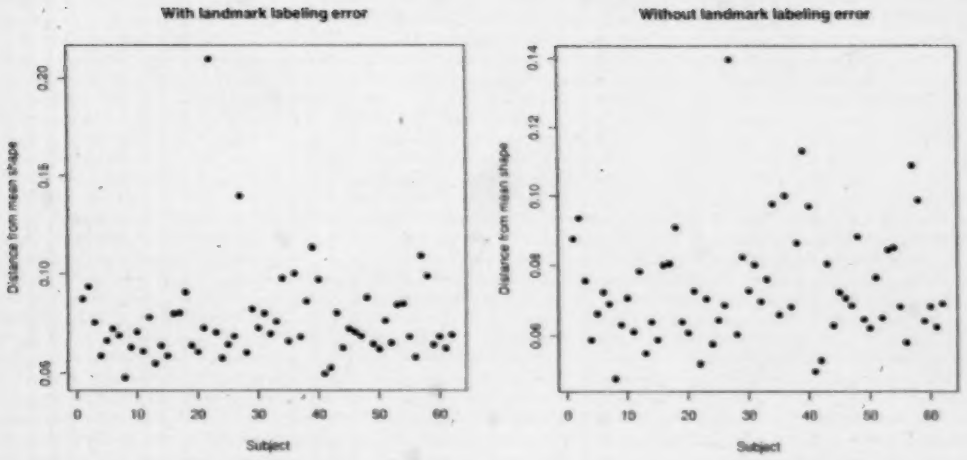


Figure 3.2: Lengths of tangent space vectors. In the left plot, the outlier corresponds to a configuration with mislabeled landmarks. The right plot shows the lengths after the mislabeling was corrected.

See Figure 3.2 for the lengths of the tangent space vectors before and after this mislabeling was corrected. Note that the moderate outlier seen at Subject 27 is due to 'true' shape differences between that configuration and the mean.

After superimposition through GPA, the mean shape was calculated for each of the three groups. They are displayed from two angles in Figure 3.3. The mean shapes appear to be rather similar, with many of the corresponding landmarks lying almost directly on top of one another. Hotelling's $T^2$ can be used for pairwise statistical comparisons between the groups.

Using the Hotelling's $T^2$ method described in Chapter 1, on p.15, the three groups are compared two at a time. During the description of Hotelling's $T^2$ test in Section 1.1.5, it was discussed that in general, the covariance matrix obtained from tangent space coordinates will not be of full rank. Given the singular covariance matrix, it is difficult to test for multivariate normality of the tangent space vectors, but with multivariate normality assumed, the p-values are given using the F distribution described on p.1.1.5. For landmark data in two-dimensions, a non-parametric method is available in Dryden's R-code *Shapes* package. The results are displayed in Table 3.1. All of the results

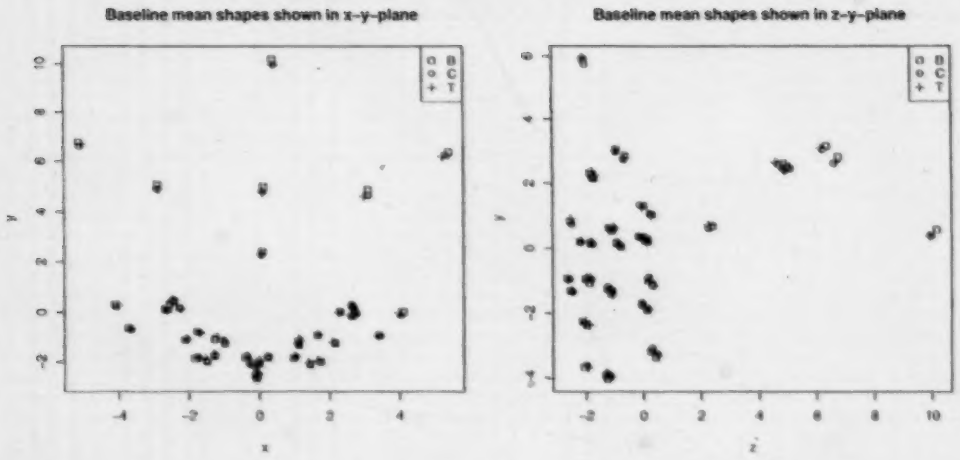


Figure 3.3: Views in the x-y and z-y planes of the three mean shapes at baseline.

| Compare groups: | B1 and C1 | B1 and T1 | C1 and T1 |
|---|---|---|---|
| F-statistic | 0.025262 | 0.029801 | 0.028554 |
| p-value | 0.999998 | 0.999999 | 0.999999 |

Table 3.1: F-statistics and p-values for Hotelling's $T^2$ test performed pairwise on the groups at baseline.

are definitely non-significant. The F-statistics are quite close to zero, and there is no evidence to reject the null hypothesis that the groups have equal mean shapes.

Again due to the singular covariance matrix, regular analysis of variance cannot be performed on the raw tangent space coordinates. Instead, a MANOVA will be performed using the first few principal components (PCs) of the tangent space data (using only B1, C1 and T1). The proportion of variance explained by the first few PCs is shown in Table 3.2. There are a few methods to decide the most appropriate number of principal components to use:

- A *scree plot* shows the proportion of total variance explained by each principal component. The scree plot for the first 20 PCs is shown in Figure 3.1.1. There will often be an 'elbow' in the plot, which delineates principal

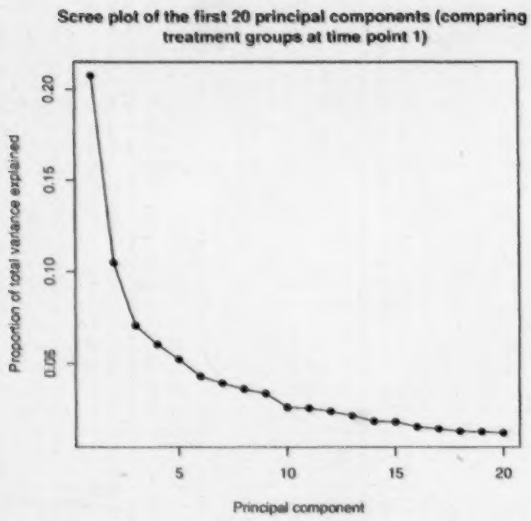


Figure 3.4: A scree plot showing the proportion of total variance explained by each of the first 20 principal components

components that account for a large proportion of variance from those that account for a relatively small amount. In the plot in Figure 3.1.1, there is no clear 'elbow', but the scree plot method would indicate the use of 1, 2, 5, or possibly 9 principal components.

- Another method of determining an appropriate number of principal components, is to use the number that cumulatively explain a certain percentage of total variance. For example, for this data set 50% of the total variance is explained by the first 6 PCs, 80% by the first 17 PCs, and 90% by the first 26 PCs.

- The method which will be used here chooses PCs that each explain at least a certain percentage of total variance. For example, PCs that each explain at least 10%, at least 5%, or at least 2% could be chosen. We will perform both an ANOVA using only the first PC, as well as a MANOVA on those PCs each accounting for 5% or more of the total variance.

Only the first five principal components each contribute at least 5% of

| | PC1 | PC2 | PC3 | PC4 | PC5 |
|---|---|---|---|---|---|
| Std dev | 0.03470724 | 0.02465053 | 0.02027949 | 0.01871909 | 0.01738854 |
| Prop. of Var. | 0.20769 | 0.10477 | 0.07091 | 0.06042 | 0.05213 |
| Cum. Prop. | 0.20769 | 0.31246 | 0.38337 | 0.44379 | 0.49592 |

Table 3.2: Standard deviation, proportion of variance, and cumulative variance of the first five principal components for the tangent space data at baseline.

| Compare groups: | B1 and C1 | B1 and T1 | C1 and T1 |
|---|---|---|---|
| T-statistic | 1.207398 | 1.300502 | 1.441240 |
| 95% CI lower bound | 1.161677(B) | 1.173055(B) | 1.161748(C) |
| 95% CI upper bound | 1.698952(B) | 1.520919(B) | 1.562171(C) |
| 95% CI lower bound | 1.164969(C) | 1.143766(T) | 1.165060(T) |
| 95% CI upper bound | 1.625738(C) | 1.541813(T) | 1.591948(T) |

Table 3.3: EDMA T statistics to compare the shape of baseline groups. See the table below for corresponding bootstrapped distributions. For 95% confidence intervals, the group used as baseline is in brackets.

the total variation. Using a univariate ANOVA on the first PC to compare the groups at baseline gives an F-statistic of 0.1542 (p-value 0.8574), and running MANOVA on the first five PCs gives an F-statistic of 0.9724 (p-value 0.4715). In both cases, no significant differences are found between the treatment groups at baseline. The ANOVA/MANOVA tables are presented in Appendix A, Tables A.6 and A.7.

### 3.1.2 EDMA methods

Next, EDMA is applied to confirm the conclusions reached in the previous section. EDMA-I, which is described on page 24, is used to avoid the assumption of a Gaussian perturbation model that is required for EDMA-II. Because EDMA-I is not a symmetric test, it will be performed twice for each pairwise group comparison, switching the group used as baseline between tests.

The observed T-statistics, along with the bootstrapped distributions are given in Table 3.3, and Figure 3.5.

None of the observed T statistics lie in the extreme tails of the bootstrapped

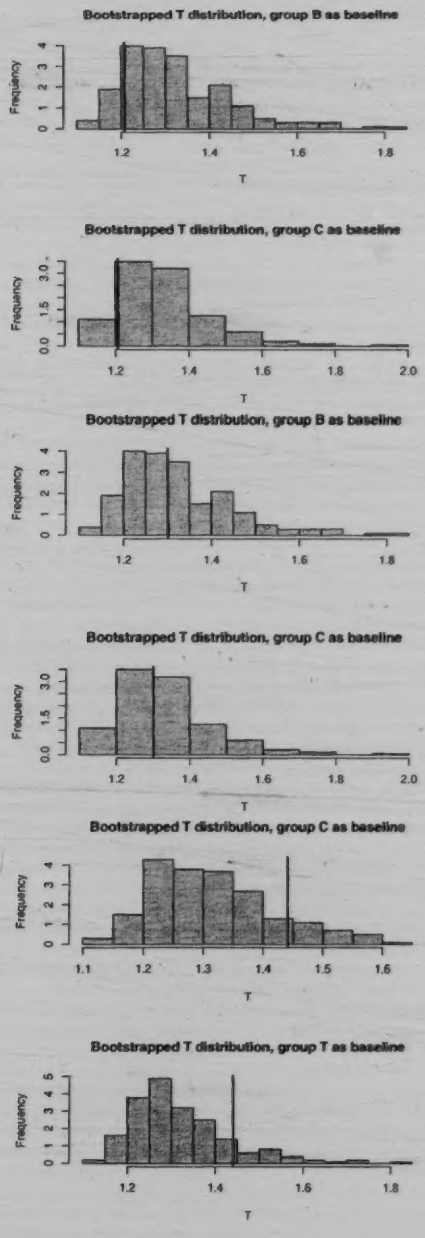


Figure 3.5: Bootstrapped T distributions using $B = 200$, comparing groups B and C (top two), B and T (middle two), and C and T (bottom two). Observed T statistic shown as bold vertical line.

distributions. This means that there is no evidence that the mean shapes differ significantly between any of the three groups at baseline. This agrees with the conclusions reached using Dryden and Mardia methods in the previous section.

## 3.2 Shape change within groups over time

Now that is has been determined that the three treatment groups did not differ significantly in mean shape before treatment began, shape analysis methods will be employed to analyze whether shape change occurs over time within each of the groups. The same subjects are followed throughout the course of treatment, so the measurements of one treatment group across multiple time points are not independent. For example, even though the groups B1, C1, and T1 are independent, the groups B1, B2, B3 and B4 are all correlated. Because of this, it is not possible to use Hotelling's $T^2$, or (M)ANOVA to analyze treatment effects of one group across time, since these methods assume independence. It is possible however, to compare the treatment effects of the three groups simultaneously by using a repeated measures ANOVA or MANOVA on the first few principal components, which is discussed in section 3.4.

EDMA does not require any assumptions about the distributions or independence the groups, so it can be applied to pairwise combinations of each group across time. The 15 possible pairwise combinations are: B2-B1, B3-B1, B4-B1, B3-B2, B4-B2, B4-B3, C3-C1, C4-C1, C4-C3, T2-T1, T3-T1, T4-T1, T3-T2, T4-T2, and T4-T3. Recall that measurements for the control group were not obtained at time point 2 (since it was assumed no change would occur in the control group).

For each of the fifteen pairwise comparisons the T statistic $T_{obs}$ is calculated using the estimated mean form difference matrix, and a bootstrap method is performed (using the earlier of the two time points as baseline) to obtain 95% confidence intervals for the test statistics. No significant differences are found between the mean shapes across time for groups B and C. For group

| | T statistic | 95% CI lower bound | 95% CI Upper Bound |
|---|---|---|---|
| B2 and B1 | 1.476944 | 1.153566 | 1.538700 |
| B3 and B1 | 1.457055 | 1.150253 | 1.584906 |
| B4 and B1 | 1.449564 | 1.139684 | 1.586500 |
| B3 and B2 | 1.427163 | 1.174960 | 2.442772 |
| B4 and B2 | 1.48592 | 1.188853 | 2.773085 |
| B4 and B3 | 1.220171 | 1.156412 | 1.676185 |
| C3 and C1 | 1.315932 | 1.179160 | 1.675546 |
| C4 and C1 | 1.370451 | 1.163255 | 1.603597 |
| C4 and C3 | 1.175318 | 1.145213 | 1.625717 |
| T2 and T1 | **1.663264** | 1.165690 | 1.548398 |
| T3 and T1 | **1.624671** | 1.144609 | 1.592262 |
| T4 and T1 | **1.702729** | 1.175452 | 1.543769 |
| T3 and T2 | 1.423590 | 1.165959 | 1.503989 |
| T4 and T2 | **1.591574** | 1.153570 | 1.571423 |
| T4 and T3 | 1.304795 | 1.160540 | 1.725676 |

Table 3.4: T statistic and 95% confidence intervals for all within-group pairwise combinations across time. The earlier time point was always chosen as baseline (for confidence intervals and for test statistics).

T however, all three of the time points after treatment began are found to have mean shapes that are significantly different than the mean shape before treatment. Also, significant differences were found in mean shape between time points 2 and 4 for group T. Test statistics and 95% confidence intervals are displayed Table 3.4.

When significant differences are detected, further tests and analyses must be performed to determine where on the object the differences are localized. One method to try and pinpoint the areas of shape difference is to analyze the bootstrapped entry-wise confidence intervals of the FDM as described on p.23. First examined was T4 vs T1, since that T statistic was the largest. Out of the $l = k(k-1)/2 = 861$ entries in the FDM, the bootstrap procedure gives 210 that have 95% confidence intervals which lie completely above 1 (i.e. in which the inter-landmark distance is larger at T4 than it was at baseline). The landmarks involved in the most of those inter-landmark distances are landmarks 13, 14, 36, 26, 37, and 28 (in that order). These landmarks are

the Zm (left and right), MB 16 Apex, upper first molar right, B 14 Apex, and upper first premolar right. Only six inter-landmark distances had 95% CIs that lay entirely below 1 (which could be expected due to chance).

The results obtained from the bootstrapped FDM for the three other significant T statistics were:

T2 vs T1: No inter-landmark distances had confidence intervals which fell completely below 1. There were 183 completely above 1, and of those, the most frequently occurring landmarks were 26, 14, 25, 20, 28 and 13.

T3 vs. T1: Only one confidence interval completely below 1: 19 (Ekm left) to 33 (MB 26 Apex). Of the 213 completely above 1, the landmarks involved the most often were the same six as in T4 vs. T1 (13, 14, 36, 26, 37, and 27).

T4 vs. T2: Between T2 and T4 there are 25 inter-landmark distances with confidence intervals completely below 1 (i.e. the distances are smaller at T4 than at T2). Out of these, landmark 25 (Ekm right) is by far the most frequent. Only three inter-landmark distances have CIs completely above 1. This may indicate some regression after removal of the appliance.

## 3.3 Shape differences between groups at times 2, 3, and 4

For each of the three remaining time points, the treatment groups are compared for significant shape differences. For each, an ANOVA is performed using the first principal component. Boxplots confirming the assumptions of normality and equal variance are shown in Figure A. None of the time points are found to have statistically significant differences between the groups, with p-values of 0.8525, 0.9392, and 0.9874 for time points 2, 3, and 4 respectively. A MANOVA is performed on those PCs whose proportion of total variance is

greater than 5%, this is the first 5 PCs are chosen for time point 2, and the first 4 are chosen for time points 3 and 4. These tests do not find statistically significant differences either (with p-values 0.0948, 0.2839, and 0.2311, respectively). The results from the ANOVA and MANOVA are displayed in Tables A.8 to A.13.

It must be noted that the choice of the number of principal components used can have effects on conclusions about statistical significance. For all three of the time points, it is possible to choose a number of principal components which give a MANOVA with a p-value below 0.05. For example, at time point 2 if the first $n$ PCs are used for a MANOVA, with $n = 8, \ldots, 18$, then a significant p-value is obtained (but if $n > 18$ the p-value will be vary between being above 0.05 and below 0.05, depending how many PCs are chosen). Similar conflicting results are seen for time points 3 and 4. Depending on how many principal components are chosen, results may be significant or not. This illustrates that care must be taken when choosing the number of PCs for analysis, and since results may be dependent on the number of PCs chosen (which is a somewhat subjective process), they must be taken 'with a grain of salt'. It is not appropriate for a researcher to choose the 'right' number of PCs based on which number gives statistically significant results. They should be chosen using a method such as a scree plot, or percent of total variation (as outlined in Section 3.1).

The groups were also analyzed for shape differences at each time point using EDMA. For each time point the three comparisons B-C, T-C and B-T were made. For the B-C and T-C comparisons, the control group was used as baseline when the bootstrap was performed (to obtain a distribution of the T statistic), but for the B-T comparison the bootstrap was performed twice, since there was no clear choice of baseline group. All of the T statistics lie within the bootstrapped 95% confidence intervals, so for all of the time points no significant differences in shape are detected between any of the treatment groups. The results are displayed in Table 3.3. Note that when group B2 is used as a baseline, the confidence interval is much wider than all the others.

| | T statistic | 95% CI lower bound | 95% CI upper bound |
|---|---|---|---|
| B2 and C2 | 1.418469 | 1.155357 | 1.617382 |
| B2 and T2 | 1.327263 | 1.166714 | 1.524947 |
| T2 and B2 | 1.327263 | 1.182718 | 3.115804 |
| T2 and C2 | 1.498158 | 1.167308 | 1.664624 |
| B3 and C3 | 1.338226 | 1.169954 | 1.596026 |
| B3 and T3 | 1.350373 | 1.173385 | 1.605891 |
| T3 and B3 | 1.350373 | 1.161546 | 1.663597 |
| T3 and C3 | 1.391712 | 1.174486 | 1.672330 |
| B4 and C4 | 1.28259 | 1.176160 | 1.609461 |
| B4 and T4 | 1.386536 | 1.151081 | 1.574167 |
| T4 and B4 | 1.386536 | 1.146254 | 1.639620 |
| T4 and C4 | 1.325425 | 1.164059 | 1.582465 |

Table 3.5: T statistics and 95% confidence intervals comparing treatment groups within each of time points 2, 3, and 4

This could indicate that group B2 displays more variability between subjects than the other groups (possibly due to just one or two subjects whose shape is quite different from the others). This large variability in group B2 is also displayed in Table 3.4. Such large variability would make it more difficult to detect small differences in shape.

Next, repeated measures MANOVA will be performed on the first few principal components derived from the tangent space coordinates, to determine if the groups change over time in significantly different ways.

## 3.4 Longitudinal shape analysis using repeated measures MANOVA

For this section the tangent space coordinates are taken for the tangent space with the pooled mean shape (over all subjects at all time points) as its pole. For the principal components calculated using these tangent space coordinates, the first four each account for at least 5% of the total variance, so they are used in the repeated measures MANOVA.

When the overall multivariate test is considered, treatment group is not found to be significant as a between-subjects variable. When considering within-subject variables, time is found to be very significant ($p \sim 0.000$), and time*group interaction is also significant ($p < 0.01$). This means that when considering the group as a whole (all 62 subjects), the scores for the first four principal components do changes across time, and that additionally, the way they change is not the same for all three groups. Further univariate tests must be performed to see which of the four principal components are changing.

Univariate tests show that the time variable is significant for the 1st, 3rd and 4th PCs (with p-values below 0.01), and that the time*group interaction is significant only for the 3rd and 4th PCs (p-values around 0.001), but not the 1st and 2nd (p-values around 0.25). Profile plots of the estimated marginal means were obtained for each principal component to display the differences between the groups through time, they are displayed in Figure 3.4.

The advantage of using principal components is in dimensionality-reduction, particularly in cases such as this where regular multivariate statistics are unable to be performed on the raw data. The disadvantage of using principal components lies in ease of interpretation. Sometimes the first one or two PCs are easily seen to correspond to certain characteristics (eg. size, or some prominent feature), but it is often difficult to identify what higher PCs 'represent'.

## 3.5 How to proceed

EDMA found statistically significant differences in shape through the course of treatment T, and repeated measures MANOVA on the first four principal components found significant time effects, and a time*group interaction. When analyzed individually however, no treatment changes were seen in group B.

Now that shape differences has been established, it is of interest where the shape differences are localized. One method to determine this is the landmark deletion method, originally discussed in Section 1.2.4. This, along with the results from the bootstrapped FDMs, can help to detect influential landmarks

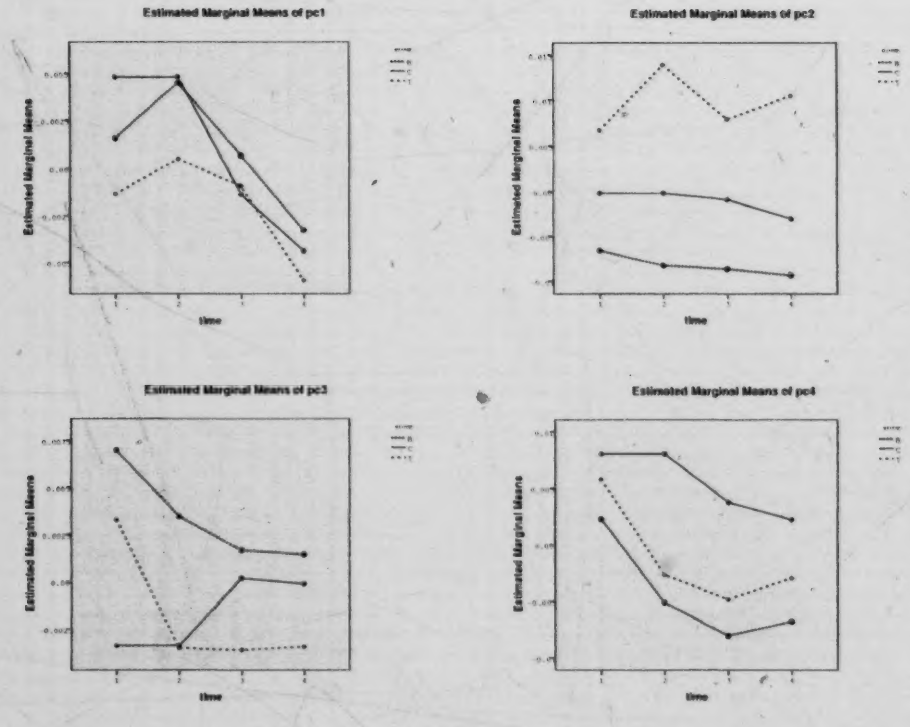
Estimated Marginal Means of pc1
Estimated Marginal Means of pc2
Estimated Marginal Means of pc3
Estimated Marginal Means of pc4
Estimated Marginal Means
time

Figure 3.6: Profile plots for the first four principal components, comparing treatment groups across time. The control group is shown as a solid line, group B as a dotted line, and group T as a dashed line.

(i.e. those landmarks that lie in areas of maximal shape difference). There is also another advantage to determining influential landmarks: if the shape analysis techniques are performed again using only the most influential landmarks, they are more likely to detect shape differences (since the variability due to shape change will represent a higher proportion of the total variance).

Detection of influential landmarks is the subject of the next chapter. Influential landmarks will represent areas of shape difference. If any landmarks are found to be particularly useful in detecting shape differences, then the shape analysis methods from this chapter will be re-run, using only the influential landmarks in the configuration.

# Chapter 4

# Finding Influential Landmarks

In the previous chapter, it was determined that some significant differences in shape exist between the treatment groups across time, but it is not yet known where these differences are localized. A bootstrap procedure performed on the FDM matrices comparing time points for treatment T found an expansion occurring (specifically involving landmarks 13, 14, 20, 25, 26, and 28) during the course of treatment, but also a slight contraction from time point 2 to time point 4.

In this chapter, the EDMA-based method of detecting influential landmarks (presented in Section 1.2.4) will be used to determine which landmarks are 'best' at distinguishing shape differences between the groups and within groups over time. All of the EDMA-based procedures employed in this text were implemented in the statistical computing language R, using code written by Jennifer Gamble (see Appendix B).

## 4.1 Mean forms within groups across time

The combinations B1, B2, B3, B4, C1, C2, C3, C4, T1, T2, T3, and T4 correspond to the treatment group and time point combinations, as indicated in the previous chapter. There are fifteen possible pairwise combinations to compare mean shapes within groups across time. For each of these comparisons

| | Influential landmarks |
|---|---|
| B2 and B1 | **32**, 31, 37, 38, 6 |
| B3 and B1 | 32, 31 |
| B4 and B1 | 36, 35, 12, 8 |
| B3 and B2 | 38, 36, 25, 13, 41, 22, 17, 21 |
| B4 and B2 | **36, 25**, 12, 8, 13, 17 |
| B4 and B3 | 37, 38 |
| C3 and C1 | **25**, 36, 14 |
| C4 and C1 | **25, 36** |
| C4 and C3 | 8, 37, 36, 31 |
| T2 and T1 | **32, 31**, 43, 8 |
| T3 and T1 | **32, 31, 33, 19** |
| T4 and T1 | **8, 32, 25, 36** |
| T3 and T2 | 19, 33 |
| T4 and T2 | 12, 8, 25, 36 |
| T4 and T3 | 8, 12, 43 |

Table 4.1: Influential landmarks as detected by the landmark deletion approach. Those shown all lowered the T statistic by at least 0.01, and those in bold lowered it by more than 0.1.

the T statistic $T_{obs}$ is calculated using the estimated mean form difference matrix. To determine the relative influence of each landmark, the T statistic is re-calculated with the $i$th landmark deleted to obtain $T_{(-i)}$ for $i = 1, \ldots, k$. The difference between $T_{obs}$ and $T_{(-i)}$ is used to rank the $k$ landmarks in order of influence. The raw $T_{(-i)}$ values are available in Tables A.14 to A.16. The landmarks whose deletion lowers the T statistic by more than 0.1 will be called the *most influential* landmarks. Those whose deletion lowers the T statistic by between 0.01 and 0.1 will be called *somewhat influential*. Table 4.1 shows the influential landmarks detected using the landmark deletion approach (in decreasing order across each row, with the most influential landmarks in bold).

The landmarks which appear in bold at least once are landmarks 8, 19, 25, 31, 32, 33 and 36. These landmarks are point A, Ekm left and right, incisal apex left and right, MB 26 Apex and MB 16 Apex.

The within-group shape analysis methods from the previous chapter can be applied again, using only this set of the seven most influential landmarks.

| | T statistic | 95% CI lower bound | 95% CI Upper Bound |
|---|---|---|---|
| B2 and B1 | **1.764662** | 1.066847 | 1.546767 |
| B3 and B1 | 1.508304 | 1.061946 | 1.493557 |
| B4 and B1 | 1.548551 | 1.059277 | 1.556786 |
| B3 and B2 | 1.979456 | 1.076727 | 3.514947 |
| B4 and B2 | 2.200955 | 1.089080 | 3.109217 |
| B4 and B3 | 1.15577 | 1.071173 | 1.620381 |
| C3 and C1 | 1.408732 | 1.080931 | 1.422927 |
| C4 and C1 | **1.562709** | 1.071991 | 1.373707 |
| C4 and C3 | 1.109301 | 1.094057 | 1.752811 |
| T2 and T1 | **1.582332** | 1.077858 | 1.525125 |
| T3 and T1 | **1.59896** | 1.099342 | 1.529163 |
| T4 and T1 | **1.736847** | 1.083779 | 1.511455 |
| T3 and T2 | 1.368207 | 1.092544 | 1.535286 |
| T4 and T2 | 1.367964 | 1.078441 | 1.510802 |
| T4 and T3 | 1.223203 | 1.104472 | 1.616627 |

Table 4.2: T statistic and 95% confidence intervals for all within-group pairwise combinations across time, using only the seven most influential landmarks. Test statistics falling outside the confidence interval are displayed in bold.

Test statistics and 95% confidence intervals comparing treatment groups across time are displayed in Table 4.1. Significant differences are again seen in group T between baseline and each of the three subsequent time points (although not between time points 2 and 4). Additionally significant differences are seen in group B between baseline and time point 2. Somewhat concerningly, significant differences are seen in the control group between the first and last time points. Possibly further analysis could reveal an outlier at one of these time points to account for such differences.

## 4.2 Mean forms across treatment groups for given time points

A similar procedure to determine influential landmarks was then performed when comparing treatments within given time points. The observed and land-

| | Influential landmarks |
|---|---|
| B2 and C2 | 31, 32 |
| B2 and T2 | 26, 27, 29, 30 |
| T2 and C2 | **32, 31** |
| B3 and C3 | 32, 31 |
| B3 and T3 | 29, 28, 33, 19 |
| T3 and C3 | **33, 19**, 28, 29 |
| B4 and C4 | 31, 29, 30, 32 |
| B4 and T4 | **29**, 33, 19, 30 |
| T4 and C4 | 19, 33, 34, 35 |

Table 4.3: T statistics and 95% confidence intervals comparing treatment groups within each of time points 2, 3, and 4

| | Pillai | approx F | num Df | den Df | Pr(>F) |
|---|---|---|---|---|---|
| Time point 1 | 0.15387 | 0.76400 | 12 | 110 | 0.6858 |
| Time point 2 | 0.3124 | 1.6969 | 12 | 110 | 0.07692 |
| Time point 3 | 0.28507 | 1.52377 | 12 | 110 | 0.1262 |
| Time point 4 | 0.21757 | 1.11890 | 12 | 110 | 0.3523 |

Table 4.4: MANOVA table using first six principal components (obtained from the tangent space coordinates using only the five most influential landmarks), comparing treatment groups within each of the four time points.

mark deleted T statistics are given in Tables A.17 to A.19. The most influential (in bold) and somewhat influential landmarks are given in Table 4.2. The landmarks whose deletion result in the T statistic decreasing by more than 0.1 are landmarks 19, 29, 31, 32, and 33. These landmarks are the Ekm left, upper canine right, incisal apex left and right, and MB 26 apex.

Using these five most influential landmarks the between-group shape analysis methods from Chapter 3 were performed once again. The first six principal components (obtained from the tangent space coordinates) each account for at least 5% of the total variance, and are used in MANOVA tests for differences between treatment groups within time points. The results are displayed in Table 4.4. Even using the most influential landmarks, there are still no statistically significant differences in shape found.

## 4.3 Repeated Measures MANOVA

Using only the seven landmarks from Section 4.1 (which include four of the five landmarks from the previous section), a repeated measures MANOVA was performed using the first five principal components from the tangent space data. For the multivariate tests, the between-subjects variable of *group* (i.e. B, C or T) was not found to be significant. The within-subjects variable *time* was found to be very significant (p-value $\sim 0$), but the interaction *group*time* failed to reach significance (p-value between 0.05 and 0.1). Further examination of the univariate tests found *time* to be significant for all but the first PC (and *group*time* significant for none of the PCs).

When using these landmarks in the above two sections, statistically significant differences were found across time, but not between groups (for given time points), so it is not surprising that no significant *group*time* interaction was found.

## 4.4 Conclusions

The results obtained when using only the most influential landmarks do not seem to be more significant than those obtained using the full landmark configurations. The main differences was that significance was found between groups B1 and B2 when using the most influential landmarks.

Overall, it appears that treatment effects are most readily seen in group T. A trend of expansion throughout the course of treatment was seen, but with some slight contraction possible between time points 2 and 4. This movement was most apparent around the areas of landmarks A, the Ekm's, the incisal apexes, MB 16 and 26, and the right first molar and pre-molar.

The fact that treatment effects were not generally seen in group B could be due to the fact that there was more natural variability in that group, which could have obscured the significance of small shape changes.

# Chapter 5

# Persistent Homology

In the previous chapters, the statistical shape analysis methods of Dryden and Mardia [15] were discussed and applied to an example data set. The implementation of these methods begins with $n$ objects of interest, each represented as a $k \times m$ landmark configuration matrix. Each configuration is transformed into a single point in some high-dimensional shape space and, if appropriate, a tangent space approximation is taken. Statistical techniques to compare groups may be employed directly in the tangent space. The points (which each correspond to a landmark configuration of interest) are taken to be samples from some multivariate distribution(s). Because of the unknown covariance structure of such distributions, inferences about the difference between mean shapes of multiple groups can be difficult.

The setting is then this: a number of points are available in a high-dimensional space. The relationship between the points of interest, particularly whether any subgroups are formed, of points corresponding to 'similar' shapes.

Persistent homology is a technique from computational algebraic topology which is used as a method of topological feature-detection when considering data with an unknown amount of 'noise'. This chapter will first discuss the background and theory of persistent homology, and then apply it to the set of tangent space points discussed previously.

## 5.1 Background and Theory

### 5.1.1 History

The ideas behind persistent homology were independently pursued by three separate groups. Frosini, Ferri, and colleagues framed their work in terms of size theory [18], [9], Robins defined persistence using an inverse systems approach, and studied fractal sets using alpha shapes [40], and Edelsbrunner, Letscher, and Zomorodian independently defined persistent homology, along with the first fast algorithms for implementation [17]. Details of persistent homology methods are presented fully in the book *Topology for Computing* by Afra Zomorodian [49], with generalized versions developed by Carlsson and Zomorodian discussed in [11] and [48]. Recent surveys of persistent homology are available in [16] and [19]

The preliminary theory and description of persistent homology given here follow the definitions and exposition as presented in Zomorodian's book [49]. Some background on topological spaces, filtrations, and homology will be required before proceeding.

### 5.1.2 Topological spaces

A *space* can vaguely be thought of as a set of points with some type of additional structure or properties. A *topology* is one such structure, which can be intuitively thought of as information about which points are 'close' to each other (i.e. information about the *connectivity* of the space). Formally, a topology must satisfy certain properties.

**Definition 5.1.** Let $X$ be a space, and $2^X$ be the set of all subsets of $X$. Then the set $T \subseteq 2^X$ is a *topology* on $X$ if

a) If $S_1, S_2 \in T$ then $S_1 \bigcap S_2 \in T$.

b) If $\{S_j | j \in J\} \subseteq T$, then $\bigcup_{j \in J} S_j \in T$.

c) $\emptyset, X \in T$.

The pair $(X, T)$ is then called a *topological space.*

From this point on, the term 'space' will be taken to mean topological space. The most familiar example to most readers of a topological space would be a *metric space.*

**Definition 5.2.** A space $X$ with distance function or *metric* $d$ satisfying

a) $d(x, y) \geq 0$ for all $x, y \in X$.

b) $d(x, y) = 0$ implies $x = y$.

c) $d(x, y) = d(y, x)$ for all $x, y \in X$.

d) $d(x, y) + d(y, z) \geq d(x, z)$ for all $x, y, z \in X$.

is called a *metric space.*

A metric space has an induced (metric) topology from $d$: using the open balls $B(x, r) = \{y \in X | d(x, y) < r\}$ with $x \in X$, and $r > 0$, the *metric topology* on $X$ consists of the set of all unions of open balls. All the spaces analyzed here will be metric spaces.

The topology of a space refers to its connectivity (and more), but not necessarily its specific geometric representation. Spaces are studied and compared through the use of *invariants.* Invariants are some fixed (i.e. *intrinsic*) properties of a space that don't change after some specific type of transformation has been applied. In the previous chapters on shape analysis, we studied shape properties which are invariant under similarity transformations. In topology, properties are studied that are invariant under *homeomorphisms* (functions that are $1-1$, onto, and continuous with continuous inverse). Invariant properties are not the only ones of interest, however. Although a topological space does exist as an independent entity, it is usually regarded as a subset of some larger space. Properties involving the way that the topological space is sitting in (i.e. *embedded* in) the larger space are called *extrinsic.*

Some examples will illustrate the concepts of the previous paragraph:

- A way to compare the connectivity of two spaces is to cut them both, and examine the new connectivity. For example, if a string is cut, then two strings are obtained, if a loop is cut, then one string is obtained. This means that the original string and the loop had different connectivities.

- Intrinsically, a loop is the same as a loop with a knot in it (in that both consist of a single, unbroken 'rope'), but in three-dimensional space if it is impossible remove the knottedness through twists and turns (without cutting it), then the extrinsic properties of the two (in three-dimensional space) are different.

- Topological properties often studied include: tunnels, knots, number of components, and voids.

Spaces of interest could include: a three dimensional object in Euclidean space, a manifold (see Section 5.1.4) in some non-Euclidean space, or (in our applications) a subset of the shape space $\sum_m^k$ (or a tangent space approximation to it) as described in Chapter 1. Full introductions to the field of topology can be found in Hocking and Young [23] or Munkres [38].

The geometry of a space (based on its particular embedding), and its topology are, of course, linked. Real world examples often involve sampled data, and small geometric modifications can change the topology of the space of interest. These small geometric modifications may be due simply to 'noise', such as measurement error in the sampling method, or they may be due to true properties of the space of interest. To learn to differentiate between noise and true properties, the concept of a filtration must first be discussed.

### 5.1.3 Filtrations

A filtration may visualized using an analogy from Zomorodian [49]. Imagine a flood on a plain: the plain consists of hills and valleys of varying heights. Initially, before the flood, the entire plain may be viewed as one connected component (you can walk anywhere without going through water). As the

water level begins to rise, the valleys start to fill up. When viewed from above, the level of the water can be shown as *contour lines* at the water's edge. At any given stage of the flood, the plain consists of whichever areas are still above water. At certain stages, hilltops will become disconnected from each other (when it is no longer possible to walk between them), and this will change the topology of the plain. The tallest hilltops will be those which remain above-water the longest, and eventually the entire plain will be submerged. The *filtration* is viewed as the entire history of the plain, at all stages of the flood.

Any space changing over time through a one-parameter function may be regarded as a filtration. In the above example, the function was the height of the water, with respect to the features on the plain. The plain may have some small hills and valleys (ant hills, say), which appear during the filtration, and change the topology of the plain, but are not considered 'true' topological features. During the flood, at one point the ant hill will be surrounded by water (a separate component), but will soon be covered. If a particular water-level during the flood is chosen as a representation of the plain, then it may be difficult to differentiate noise from true properties. Instead, if the entire filtration is considered, then properties that last a long time throughout the flood can be considered true properties, while those that appear and disappear quickly are considered noise. *Persistence* is a measure of topological attributes, which ranks them by their lifetime in a filtration. The assumption here is that lifetime is related to importance.

It is this method through which the geometry and topology of a space are considered simultaneously. For a given space of interest, a filtration is chosen based on its geometry. Topological features of the space are then measured throughout the filtration, and the persistent features are considered to be true features of the space, while others are considered noise (possibly due to sampling error, etc). Note that the conclusions about which features are topologically significant is not invariant to the choice of filtration. Returning to the flooding plain example: imagine the same plain, but tilted so that one

end was higher than the other. In this case the filtration would fill up valleys at the lower end of the plain first (perhaps even covering large hills at the lower end before reaching valleys at the top end). Conclusions drawn about persistent features (i.e. big hills) using this filtration would likely be different than those using the original (un-tilted) filtration.

### 5.1.4 Simplicial Complexes

A *manifold* may be intuitively thought of as a topological space that locally looks like $\mathbb{R}^n$. Its global structure may be far more complicated, however. See any topology textbook (such as [38]) for a full mathematical definition. Common examples of manifolds include the sphere or torus embedded in three-space, or a non-orientable surface such as a Mobius strip.

For computational purposes, surfaces need to be represented in a finite way, so are often sampled and represented by triangulations. A triangulation is a simplicial complex, which allows representation of a (continuous) surface in a finite (and thus computable) way. We begin with some necessary definitions.

**Definition 5.3 (combinations).** For a set of points $S = \{p_0, p_1, \ldots, p_k\}$ in $\mathbb{R}^d$. A *linear combination* is $x = \sum_{i=0}^{k} \lambda_i p_i$, for some $\lambda_i \in \mathbb{R}$, an *affine combination* is a linear combination with $\sum_{i=0}^{k} \lambda_i = 1$. A *convex combination* is an affine combination with $\lambda_i > 0$, for all $i$. The set of all convex combinations is the *convex hull.*

**Definition 5.4 (independence).** If no point in $S$ is a linear (affine) combination of the other points in $S$, then $S$ is called *linearly (affinely) independent.*

**Definition 5.5 ($k$-simplex).** A *$k$-simplex* is the convex hull of $k+1$ affinely independent points $S = \{v_0, v_1, \ldots, v_k\}$. The points in $S$ are called the *vertices* of the simplex.

A $k$-simplex, $\sigma$, is a $k$-dimensional subspace of $\mathbb{R}^d$. Write $\dim\sigma = k$.

**Definition 5.6 (face, coface).** For $k$-simplex $\sigma$ with vertices $S = \{v_0, v_1, \ldots, v_k\}$, let $\tau$ be defined by $T \subseteq S$. Then $\tau$ is called a *face* of $\sigma$ and $\tau$ has $\sigma$ as a *coface*. Write $\sigma \geq \tau$ and $\tau \leq \sigma$. Note that $\sigma \leq \sigma$ and $\sigma \geq \sigma$.

**Definition 5.7 (simplicial complex).** A finite set of simplices $K$, that satisfy the properties

(a) $\sigma \in K, \tau \leq \sigma \Rightarrow \tau \in K$;

(b) $\sigma, \sigma' \in K \Rightarrow \sigma \bigcap \sigma' \leq \sigma, \sigma'$ or $\sigma \bigcap \sigma' = \emptyset$

is called a *simplicial complex*.

Define the *dimension* of $K$ as $\dim K = \max\{\dim \sigma | \sigma \in K\}$, and call the zero-simplicies of $K$ *vertices* or *nodes*. Defining the empty set $\emptyset$ as the (-1)-*simplex*, then a $k$-simplex has $\binom{k+1}{l+1}$ faces of dimension $l$, and

$$\sum_{l=-1}^{k} \binom{k+1}{l+1} = 2^{k+1}$$

faces in total.

**Example 5.1.** Note requirement (a) in the above definition. If a simplex is in $K$, then all of its faces must be in $K$ also. For example, if a simplicial complex contains a 2-simplex (i.e. a triangle), then each of the three 'sides' of the triangle, and its three vertices must also be simplices in the complex.

Simplices may quickly become very large objects, so are not well suited for use in hand-computations, but the ease with which they may be defined and stored makes them ideal for use in computer algorithms.

The above definition of a simplicial complex was a geometric one, but a non-geometric abstract definition is also possible.

**Definition 5.8 (abstract simplicial complex).** For a set $K$, with $\mathcal{S} \subseteq 2^K$, if:

(a) For all $v \in K, \{v\} \in \mathcal{S}$. We call the sets $\{v\}$ the *vertices* of $K$.

(b) If $\tau \subseteq \sigma \in \mathcal{S}$, then $\tau \in \mathcal{S}$.

then $K$ is called an *abstract simplicial complex*, and the subsets of $K$ contained in $\mathcal{S}$ are called *(abstract) simplices.*

If $|\sigma| = k+1$, call $\sigma$ a *$k$-simplex of dimension $k$*. For $\tau \subseteq \sigma, \tau$ is a *face* of $\sigma$ and $\sigma$ is a *coface* of $\tau$.

**Definition 5.9 (vertex scheme).** For simplicial complex $K$ with vertices $V$, let $\mathcal{K}$ be the collection of all subsets $\{v_0, v_1, \ldots, v_k\}$ of $V$ such that the vertices $v_0, v_1, \ldots, v_k$, span a simplex of $K$. The collection $\mathcal{K}$ is called the *vertex scheme* of $K$.

**Example 5.2.** Using the simplicial complex from Example 5.1 above (call the nodes of the triangle $a, b$ and $c$), the corresponding vertex scheme is:

$$\mathcal{K} = \{\{abc\}, \{ab\}, \{ac\}, \{bc\}, \{a\}, \{b\}, \{c\}\}.$$

The collection $\mathcal{K}$ is an abstract simplicial complex. It allows us to compare simplicial complexes easily, using isomorphisms.

**Definition 5.10 (isomorphism).** For abstract simplicial complexes $K_1, K_2$ with vertex sets $V_1, V_2$, an *isomorphism* between $K_1, K_2$ is a bijection $\phi : V_1 \to V_2$, such that the sets in $K_1$ and $K_2$ are the same under the renaming of the vertices by $\phi$ and its inverse.

**Theorem 5.1.** *Every abstract complex $\mathcal{S}$ is isomorphic to the vertex scheme of some simplicial complex $K$. Two simplicial complexes are isomorphic iff their vertex schemes are isomorphic as abstract simplicial complexes.*

The proof is in *Elements of Algebraic Topology* by J. R Munkres (1984).

**Definition 5.11 (geometric realization).** For a simplicial complex $K$, if $\mathcal{S}$ is an abstract simplicial complex isomorphic to the vertex scheme of $K$, then call $K$ a *geometric realization* of $\mathcal{S}$. It is uniquely determined up to an isomorphism, linear on the simplices.

Note the distinction between an abstract complex and its representation as a simplicial complex. An abstract complex is a purely combinatorial object, which is easily stored and manipulated in a computer system. We can map the vertices of such a complex into some space where the complex is realized. These two constructions correspond to our distinction between topological and geometrical components.

While a simplex is the power set of its vertices, a simplicial complex is a combination of a number of simplices (and so is only a subset of the power set of its vertices). This subset is specified by the requirements in Definition 5.8. We will use simplicial complexes to represent manifolds.

**Definition 5.12 (underlying space).** The *underlying space* $|K|$ of a simplicial complex $K$ is $|K| = \bigcup_{\sigma \in K} \sigma$.

**Definition 5.13 (triangulation).** A *triangulation* of a topological space $\mathbf{X}$ is a simplicial complex $K$ such that $|K|$ is homeomorphic to $\mathbf{X}$.

**Example 5.3.** The boundary of a tetrahedron (3-simplex) is homeomorphic to a sphere.

Note: The term 'triangulation' here can mean complexes of any dimension. Often in computer graphics, a triangulation refers to the 2-dimensional case only, then *tetrahedralizations* are 3-dimensional, etc. Another term often used is a *mesh* which might also include other elements (such as quadrangles or cubes) than just simplices. The spaces explored here using persistent homology will be simplicial complexes. They will be built incrementally (through a filtration) in such a way that all the subsets generated are also complexes. Orientability may be defined for simplicial complexes.

**Definition 5.14 (orientation).** For a simplicial complex $K$, an *orientation* of a $k$-simplex $\sigma \in K$, $\sigma = \{v_0, v_1, \ldots, v_k\}, v_i \in K$, is an equivalence class of orderings of the vertices of $\sigma$, where

$$(v_0, v_1, \ldots, v_k) \sim (v_{\tau(0)}, v_{\tau(1)}, \ldots, v_{\tau(k)})$$

are equivalent orderings if $v_{\tau(0)}, v_{\tau(1)}, \ldots, v_{\tau(k)}$ can be obtained through an even number of transpositions of $v_0, v_1, \ldots, v_k$. We denote an *oriented simplex*, a simplex with an equivalence class of orderings, by $[\sigma]$. If $\sigma = \tau$, but $\sigma$ and $\tau$ have different orientations, then write $[\sigma] = -[\tau]$.

The concept of orientability may be extended from simplices to triangulated $d$-manifolds.

**Definition 5.15 (orientability).** Two $k$-simplices sharing a $(k-1)$-face $\sigma$ are *consistently oriented* if they induce different orientations on $\sigma$. A triangulable $d$-manifold is *orientable* if all $d$-simplices can be oriented consistently. Otherwise, the $d$-manifold is *nonorientable*.

**Definition 5.16 (subcomplex).** A simplicial complex $L \subseteq K$ is called a *subcomplex* of $K$.

**Definition 5.17 (filtration).** A *filtration* of a complex $K$ is a nested sequence of subcomplexes, $\emptyset = K^0 \subseteq K^1 \subseteq K^2 \subseteq \cdots \subseteq K^m = K$. We call a complex $K$ with a filtration a *filtered complex*.

This filtration of simplicial complexes will allow ranking of topological features by their persistence in the filtration. Before moving on to a more rigorous definition of persistence, a choice of topological classification (i.e. which topological features are deemed of interest) must be made.

### 5.1.5 Homology

There are a number of methods available to determine topological invariants. For a given choice, the set of all topological spaces will be partitioned into classes that display the same invariant properties. Two common methods of classifying topological spaces are by *topological type* or *homotopy type* (see an algebraic topology text such as [22] or [37]), but the method most appropriate for the purposes used here is *simplicial homology*.

In general, homology groups describe how cells of dimension $n$ attach to cells of dimension $n-1$. Simplicial homology specifically is a weaker form of

homology which is defined only for simplicial complexes, and is combinatorial in nature. Simplicial homology is still however, invariant to the underlying space of a simplicial complex.

The definitions of homology groups require a background in abstract algebra (specifically group theory), which will not be given here. See Dummit and Foote for full exposition, or Chapter 3 of Zomorodian [49] for a more compact summary of the tools necessary.

A basic concept in algebraic topology is that of the fundamental group. A full definition is beyond the scope of this text, but intuitively it is a description of how many ways a loop (a path beginning and ending at the same base point) can be made in the space. Loops that can be continuously deformed into each other are considered 'the same', and the torus is given as a canonical example: one loop can be made around the 'hole', and another around the 'ring'. See [22] for a full mathematical definition of the fundamental group. The fundamental group is concerned with families of maps on a surface, where these maps are paths and loops, and this construction motivates the definition of simplicial homology groups.

To define homology groups, we need simplicial analogs of paths and loops. We do this using the chain group of oriented simplices.

**Definition 5.18 (chain group).** For simplicial complex $K$, the free Abelian group on the oriented $k$-simplices is called the *kth chain group* of $K$, and is denoted $\langle C_k(K), + \rangle$. An element $\sum_q n_q[\sigma_q] \in C_k(K)$ is called a *k-chain*, where $n_q \in \mathbb{Z}, \sigma_q \in K$.

To look at the connectivity between two immediate dimensions, we define a structure-relating map between chain groups.

**Definition 5.19 (boundary homomorphism).** Let $K$ be a simplicial complex and $\sigma \in K, \sigma = [v_0, v_1, \ldots, v_k]$. The boundary homomorphism $\partial_k : C_k(K) \to C_{k-1}(K)$ is

$$\partial_k \sigma = \sum_i (-1)^i [v_0, v_1, \ldots, \hat{v}_i, \ldots, v_k]$$

where $\hat{v}_i$ indicates that $v_i$ was deleted from the sequence.

For a given orientation of a simplex, any ordering of its vertices produces the same $\partial_k$, i.e. $\partial_k$ is well-defined.

**Example 5.4 (boundaries).**

- $\partial_1[a,b] = b - a$
- $\partial_2[a,b,c] = [b,c] - [a,c] + [a,b] = [b,c] + [c,a] + [a,b]$
- $\partial_3[a,b,c,d] = [b,c,d] - [a,c,d] + [a,b,d] - [a,b,c]$

If we take the boundary of the boundary of a triangle, we get

$$\partial_1\partial_2[a,b,c] = [c] - [b] - [c] + [a] + [b] - [a] = 0$$

This is intuitively correct: the boundary of a triangle is a cycle, and a cycle does not have a boundary. In fact, this intuition generalizes to all dimensions.

**Theorem 5.2.** $\partial_{k-1}\partial_k = 0$, *for all* $k$.

The proof amounts to writing out $\partial_{k-1}\partial_k$ using Definition 5.19, then canceling terms.

The boundary homomorphism allows the $n$-dimensional complex $K$ to be understood using the *chain complex*

$$0 \longrightarrow C_n \xrightarrow{\partial_n} C_{n-1} \xrightarrow{\partial_{n-1}} \cdots \xrightarrow{\partial_2} C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0,$$

where the 0's on the right and left correspond to the facts that $\partial_0 = 0$, and $C_{n+1} = 0$ (since there are no $(n+1)$-simplices in $K$). The images and kernels of these maps are subgroups of $C_k$, and are given specific names.

**Definition 5.20.** $\ker K\partial_k := Z_k$ is called the *kth cycle group*, and is a free Abelian normal subgroup of $C_k$. Elements of $Z_k$ are called *k-cycles*.

$\text{im}\partial_{k+1} := B_k$ is called the *kth boundary group*, and is both a free Abelian normal subgroups of $C_k$, and a normal subgroup of $\ker \partial_k$. Elements of $B_k$ that

are chains are called *k-boundaries*. These boundaries are also called *bounding cycles*, while cycles that are non in $B_k$ are called *non-bounding cycles*.

Non-bounding cycles can be thought of as 'empty', since they don't bound a higher dimensional cycle. Simplicial homology groups may now be defined.

**Definition 5.21 (homology group).** The *kth homology group* is defined as

$$H_k = Z_k/B_k = \ker \partial_k / \mathrm{im} \partial_{k+1}.$$

Call $z_1$ and $z_2$ *homologous* if $z_1 = z_2 + B_k, z_1, z_2 \in Z_k$, and denote this as $z_1 \sim z_2$.

Homology groups describe spaces through their Betti numbers and torsion subgroups, although the applications here only consider $\beta_0$ so do not involve torsion.

**Definition 5.22.** For a simplicial complex $K$, its $k$th Betti number $\beta_k$ is the rank of the free part of $H_k$, and is denoted $\beta(H_k)$. Note that

$$\beta_k = \mathrm{rank} H_k = \mathrm{rank} Z_k - \mathrm{rank} B_k.$$

For a torsion-free complex $K$, Betti numbers have an intuitive meaning: $\beta_0$ gives the number of components of the complex, $\beta_1$ is the rank of the basis for *tunnels* of $K$ (i.e. for the underlying space of $K$, this is the number of cuts possible without dividing the space into separate components), and $\beta_2$ gives the number of *voids*.

**Example 5.5.** For the well-known spaces of a sphere, a torus, and a Klein bottle, all have $\beta_0 = 1$, since they consist of one connected component. Both the Klein bottle and the torus have $\beta_1 = 2$, and $\beta_2 = 1$, whereas the sphere has $\beta_1 = 0$ and $\beta_2 = 1$ (since it has no tunnels but one void).

The first few Betti numbers will be the primary topological invariants considered when analyzing a filtration of a space. As the filtration progresses,

the number of connected components, tunnels and voids will change, with the long-lasting features representing true properties of the space.

### 5.1.6 Persistent Homology

Applying homology to filtrations, cycle groups, boundary groups and homology groups (and thus Betti numbers) are calculated for each complex in the filtration.

**Definition 5.23.** For space $\mathbb{X}$ with filtration $K^l = \{K^0, K_1, \ldots, K_m\}$ (where $\emptyset = K_0 \subseteq \ldots \subseteq K^m = K$ is a nested sequence of subcomplexes, as defined in Definition 5.17), denote the cycle, boundary and homology groups using the superscript $l$. In other words, $Z_k^l = Z_k(K^l)$ and $B_k^l = B_k(K^l)$ are the $k$th cycle and boundary groups of $K^l$, and $H_k^l = H_k(K^l)$ is the $k$th homology group of $K^l$. Similarly denote the $k$th Betti number of $K^l$ as $\beta_k^l$.

Features of interest are nonbounding cycles with long lives during the filtration. These can be defined more rigorously for given persistence-length $p$, as the cycles that are nonbounding at a given point, and also nonbounding for the next $p$ complexes.

**Definition 5.24 (persistent homology).** The *p-persistent kth homology group* of filtration $K^l$ is defined as

$$H_k^{l,p} = Z_k^l / (B_k^{l+p} \bigcap Z_k^l).$$

The rank of $H_k^{l,p}$ gives the *p-persistent kth Betti number* $\beta_k^{l,p}$.

**Definition 5.25 (persistence).** A non-bounding $k$-cycle $z$ has homology class $[z] \in H_k^i$. $z$ is *created* at time $i$ by the arrival of simplex $\sigma$, and $z' \in [z]$ is turned into a boundary (i.e. *destroyed*) at time $j \geq i$ by the arrival of simplex $\tau$. So $z' \in B_k^j$, and its destruction merges $[z]$ with an older class of cycles, and decreases the rank of the homology group. The *persistence* of $z$ (and of $[z]$) is $j - i - 1$. Call $\sigma$ a *positive* simplex, and the *creator* of $[z]$. Call $\tau$ a

*negative* simplex, and the *destroyer* of $[z]$. A cycle with no destroyer is given persistence of $\infty$.

A number of methods are available to visualize the persistent features of a given filtered complex. The most common, which will be used here, is in the form of a *barcode*. For a homology class $[z]$ which is created at time $i$ and destroyed at time $j$, a horizontal line will be drawn beginning at time $i$ and ending at time $j$. For a given $H_k$, multiple homology classes may exist, with each represented as a horizontal line, staggered vertically for visualization. The number of intervals at any given point during the filtration, is the rank of $H_k$ (i.e. $\beta_k$) at that time point.

In the next section we apply the idea of persistent homology to the tangent space representation of configurations from the orthodontic data set.

## 5.2 Persistent homology applied to tangent space data

When using statistical shape analysis methods on the orthodontic data set, we last left the shapes as single points in a tangent space approximation to the shape space. Using traditional statistical analyses, we were unable to distinguish any significant differences in the mean shapes of the groups (either between treatments or across time). Here, persistent homology methods will be applied to the points to determine if they fall into any distinct subgroups. This will be done by examining the persistence of the 0th Betti number during a filtration. Distinct subgroups will appear as persistent lines on the barcode.

The filtration chosen here is a Cech filtration:

**Definition 5.26 (Cech filtration).** Let $V = \{v_0, v_1, \ldots, v_{n-1}\}$ be a set of vertices in $\mathbb{R}^n$. For $i = 1, \ldots, n$, let $b_i$ be a ball with radius $t$ centered at $v_i$. When $b_i \bigcap b_j \neq \emptyset$ the 1-simplex $\{v_i v_j\}$ joins the complex. Similarly, for $2 \leq k \leq n$ when $b_{i_1} \bigcap \ldots \bigcap b_{i_k} \neq \emptyset$ the $k$-simplex $\{v_{i_1} \ldots v_{i_k}\}$ joins the complex. At time $t = 0$, the Cech complex of $V$ consists of only the $k$ 0-dimensional

simplices (and thus has $k$ separate components). As $t$ increases, the first simplex to join the complex is the 1-simplex connecting the two closest vertices (decreasing the number of components by one). This continues until all vertices are connected, and there is just one connected component.

If there are $k$ distinct subgroups formed by the vertices, then the number of components will drop from $n$ to $k$ rather quickly, stay at $k$ for some time, and then drop relatively quickly to 1.

To apply these methods to the example data set, the tangent space approximation is taken using the grand mean (over all groups at all time points) as the pole. Then persistent homology methods may be applied to subsets of interest (corresponding to the same treatment group over time, or multiple treatment groups at a single time point) to determine if distinct groups of vertices are seen. If they are, then those vertices correspond to shapes that are more similar. Also, the relationship between the points (using all treatment groups) may be compared across time points.

A software package called PLEX has been developed in Gunnar Carlsson's applied topology research group at Stanford. It is written in MatLab by Patrick Perry and Vin de Silva, and incorporates a persistent homology C++ library written by Lutz Kettner and Afra Zomorodian. It was used to perform the persistent homology calculations, and obtain barcodes. Determining which persistent components correspond to which initial nodes was performed using code written in R by Jennifer Gamble

### 5.2.1 Results

The Cech filtrations are performed separately for each time point, and compared to see if the structure of the components changes through the course of treatment. It is found that for all four time points the persistent components do not correspond to the treatment groups. This indicates that there is overlap between the groups in the tangent space. Using multidimensional scaling to graph the points, it is seen that variability within the groups is much larger

than variability between the groups (which was also concluded in Chapter 3). The persistent components correspond to single outlying shapes. The barcode obtained at time point 3 is shown in Figure 5.1 below, along with a plot of the Euclidean distance from each point to the mean shape, and the points graphed using the first two dimensions using multidimensional scaling.

Performing persistent homology on group T across all four time points, some persistent components were seen. Upon inspection, these are clusters of four points (corresponding to a single subject across the four measurements) for subjects that are slight outliers (i.e. have rather different shape from the other subjects). So persistent homology did not distinguish between the time points, but between the subjects instead. This is due to the large inter-subject variability as compared to the variability between groups. The barcode, along with a 2d representation of the data through multidimensional scaling are given in Figure 5.2.

The results using persistent homology are consistent with some of those from previous chapters. Hotelling's $T^2$, ANOVA on the first PC, and MANOVA on the first few PCs did not detect any shape differences between the groups. Using persistent homology, there is no visible clustering of the points in the tangent space.

See the next chapter for directions for future research.

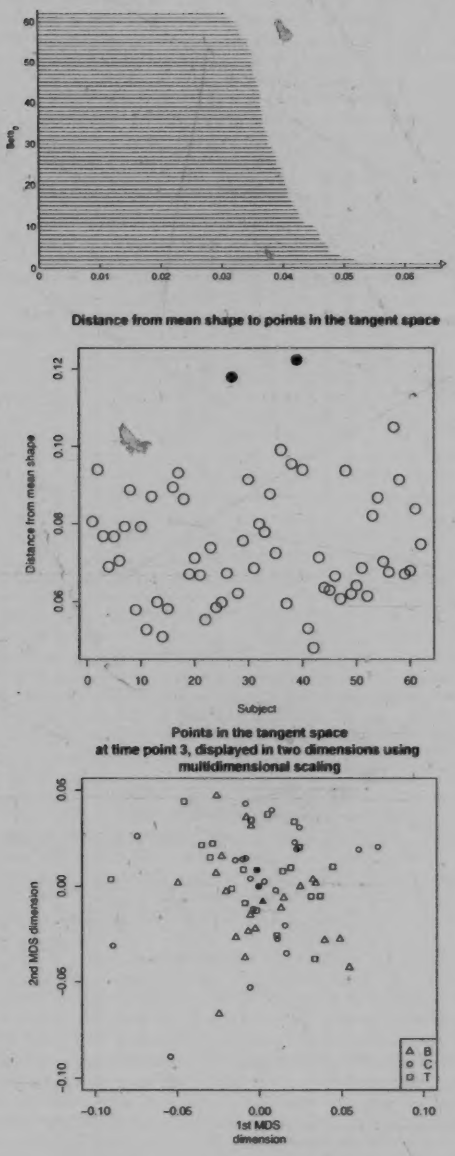


Figure 5.1: At the top is the barcode corresponding to the 0th Betti number at time point 3. It ends with one persistent component, as expected. The two persistent components which seem a bit longer than the others correspond to two single shapes which are slight outliers (and not distinct subgroups). These outliers are highlighted in a plot of the Euclidean distances from each shape to the mean shape (middle). At the bottom is a multidimensional scaling representation of the points in the tangent space, with the group means shown as filled-in points.

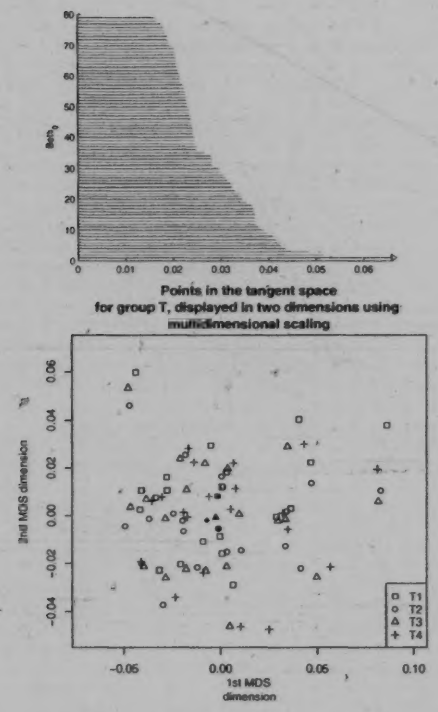


Figure 5.2: Barcode corresponding to the 0th Betti number using group T. Persisting components correspond to clusters of four points, belong to the same subject at multiple time points. On the right is a MDS representation of the points in the tangents space (within groups means filled in).

# Chapter 6

# Directions for Future Research

Regardless of the method, or shape descriptor used, an overarching goal of statistical shape analysis is to compare shapes directly in some shape space, where their relation in the shape space may be used to determine the specific way in which the shape of the (real world) objects differ. In the case of the shape space for landmark coordinate data, there is still much to learn. When all the shapes under consideration are quite similar, tangent space approximations may be appropriate, but this situation is not always the case.

One possible direction is to use a method of geodesic-estimation, such as ISOMAP instead of a tangent space approximation. Distances can be estimated using ISOMAP, and those values used to build a filtration on the points (to perform persistent homology methods). This should approximate distances in the shape space more closely than a tangent space projection, especially for shapes with larger deformations.

Another way to check for curving or non-linearity in the shape space is to use the persistent homology methods to compute higher Betti numbers for the sample data. It is not expected that there would be any tunnels or voids within the sample data, so if any were seen it would likely be an artifact of the tangent space projection. This would imply that the shape space curves back on itself in some way, resulting in nonzero higher Betti numbers in the tangent space.

There are methods developped by Carlsson, Zomorodian, Collins and Guibas [12], [14] for using the barcodes (obtained via persistent homology on 3D point cloud data) as shape descriptors. This involves the use of a metric on the barcode space to determine distances between shapes. Perhaps a metric such as this could be used to compare groups of objects in the tangent space (i.e. a way to measure the distance between the treatment group clusters). Methods applying a statistical approach to persistent homology, as in [8] could also be used in the shape analysis setting.

Statistical shape analysis is without question a field with much breadth and depth. The next step I would like to take is to apply statistical methodologies (eg. density estimation) in conjunction with persistent homology methods to study the shape space (as represented by points derived from a sample of landmark configurations).

# Bibliography


[1] Jos M. F. Ten Berge. Orthogonal procrustes rotation for two or more matrices. *Psychometrika*, 42(2):267–276, 1977.

[2] Harry Blum. A transformation for extracting new descriptors of shape. In W. Wathen-Dunn, editor, *Models for the Perception of Speech and Visual Form*. MIT Press, 1967.

[3] Fred L. Bookstein. *The Measurement of Biological Shape and Shape Change*. Springer-Verlag, 1st edition, 1978.

[4] Fred L. Bookstein. A statistical method for biological shape comparisons. *Journal of Theoretical Biology*, 107(3):475–520, 1984.

[5] Fred L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(6):567–585, 1989.

[6] Fred L. Bookstein. *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press, 1st edition, 1991.

[7] Fred L. Bookstein. A hundred years of morphometrics. *Acta Zoologica*, 44(1):7–59, 1998.

[8] Peter Bubenik and Peter Kim. A statistical approach to persistent homology. *Homology, Homotopy and Applications*, 9(2):337–362, 2007.

[9] F. Cagliari, M. Ferri, and P. Pozze. Size functions from a categorical viewpoint. *Acta Applicandae Mathematicae*, 67(3):225–235, 2001.

[10] Erik Carlsson, Gunnar Carlsson, and Vin de Silva. An algebraic topological method for feature identification. *International Journal of Computational Geometry & Applications*, 16(4):291–314, 2006.

[11] Gunnar Carlsson and Afra Zomorodian. The theory of multidimensional persistence. *Proceedings of the 23rd Annual Symposium on Computational Geometry*, 6:184–193, 2007.

[12] Gunnar Carlsson, Afra Zomorodian, Anna Collins, and Leonidas J. Guibas. Persistence barcodes for shapes. *International Journal of Shape Modeling*, 11(2):149–187, 2005.

[13] James M. Cheverud and Joan T. Richtsmeier. Finite-element scaling applied to sexual dimorphism in rhesus macaque (*Macaca Mulatta*) facial growth. *Systematic Zoology*, 35(3):381–399, 1986.

[14] Anne Collins, Afra Zomorodian, Gunnar Carlsson, and Leonidas J. Guibas. A barcode shape descriptor for curve point cloud data. *Computers & Graphics*, 28(6):881–894, 2004.

[15] Ian L. Dryden and Kanti V. Mardia. *Statistical Shape Analysis*. John Wiley and Sons, 1st edition, 1998.

[16] Herbert Edelsbrunner and John Harer. Persistent homology - a survey. *Contemporary Mathematics*, to appear.

[17] Herbert Edelsbrunner, David Letscher, and Afra Zomorodian. Topological persistence and simplification. *Discrete Computational Geometry*, 28(4):511–533, 2002.

[18] P. Frosini and C. Landi. Size theory as a topological tool for computer vision. *Pattern Theory and Image Analysis*, 9(4):596–603, 1999.

[19] Robert Ghrist. Barcodes: The persistent topology of data. *Bulletin of the American Mathematical Society*, 45(1):61–75, 2007.

[20] J. C. Gower. Generalized procrustes analysis. *Psychometrika*, 40(1):33–51, 1975.

[21] Ulf Grenander and Michael I. Miller. *Pattern Theory: From Representation to Inference*. Oxford University Press, 1st edition, 2007.

[22] Allen Hatcher. *Algebraic Topology*. Cambridge University Press, 1st edition, 2001.

[23] John G. Hocking and Gail S. Young. *Topology*. Courier Dover Publications, 1st edition, 1961.

[24] Richard A. Johnson and Dean W. Wichern. *Applied Multivariate Statistical Analysis*. Pearson Prentice Hall, 6th edition, 2007.

[25] Sarang C. Joshi and Michael I. Miller. Landmark matching via large deformation diffeomorphisms. *IEEE Transactions on Image Processing*, 9(8):1357–1370, 2000.

[26] David G. Kendall. Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, 16(2):81–121, 1984.

[27] David G. Kendall. A survey on the statistical theory of shape. *Statistical Science*, 4(2):87–99, 1989.

[28] Subhash R. Lele. Some comments on coordinate-free and scale-invariant methods in morphometrics. *American Journal of Physical Anthropology*, 85(4):407–417, 1991.

[29] Subhash R. Lele and Joan T. Richtsmeier. A coordinate-free approach for comparing biological shapes using landmark data. *American Journal of Physical Anthropology*, 86(3):415–427, 1991.

[30] Subhash R. Lele and Joan T. Richtsmeier. On comparing biological shapes: Detection of influential landmarks. *American Journal of Physical Anthropology*, 87(1):49–65, 1992.

[31] Subhash R. Lele and Joan T. Richtsmeier. *An Invariant Approach to Statistical Analysis of Shapes*. Chapman and Hall/CRC Press, 1st edition, 2001.

[32] Benjamin Franklin Lowe, Ceib Phillips, Pete E. Lestrel, and Henry W. Fields. Skeletal jaw relationships: A quantitative assessment using elliptical fourier functions. *Angle Orthodontist*, 64(4):299–310, 1994.

[33] Kenneth O. McGraw and S. P Wong. Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1):30–46, 1996.

[34] Peter W. Michor and David Mumford. An overview of the riemannian metrics on spaces of curves using the hamiltonian approach. *Applied and Computational Harmonic Analysis*, 23(1):74–113, 2007.

[35] M. I Miller and L. Younes. Group actions, homeomorphisms, and matching: A general framework. *International Journal of Computer Vision*, 41(1/2):61–84, 2001.

[36] Melvin L. Moss, Henning Vilmann, Letty Moss-Salentijn, Kasturi Sen, Hector M. Pucciarelli, and Richard Skalak. Studies on orthocephalization: Growth behaviour of the rat skull in the period 13-49 days as described by the finite element method. *American Journal of Physical Anthropology*, 72(3):323–342, 1986.

[37] James R. Munkres. *Elements of Algebraic Topology*. Westview Press, 1st edition, 1984.

[38] James R. Munkres. *Topology*. Prentice Hall, 2nd edition, 1999.

[39] Stephen M. Pizer and Kaleen Siddiqi. *Medial Representations: Mathematics, Algorithms and Applications*. Springer, 1st edition, 2008.

[40] V. Robins. Towards computing homology from finite approximations. *Topology Proceedings*, 24(2):503–524, 1999.

[41] F. James Rohlf and Leslie F. Marcus. A revolution in morphometrics. *Trends in Ecology and Evolution*, 8(4):129–132, 1993.

[42] Patrick E. Shrout and Joseph L. Fleiss. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2):420–428, 1979.

[43] D'Arcy Wentworth Thompson. *On Growth and Form*. Cambridge University Press, 1st edition, 1917.

[44] T. P Wallace and P. A. Wintz. An efficient three-dimensional aircraft recognition algorithm using fourier descriptors. *Computer Graphics and Image Processing*, 13(1):99–126, 1980.

[45] R. E White. *An Introduction to the Finite Element Method with Applications to Nonlinear Problems*. Wiley, 1st edition, 1985.

[46] Charles Zahn and Ralph Roskies. Fourier descriptors for plane closed curves. *IEEE Transactions on Computers*, 21(3):269–281, 1972.

[47] O. C. Zienkiewicz and D. W Kelly. *The Finite Element Method Vol 2: Solid and Fluid Mechanics, Dynamics, and Non-Linearity*. McGraw-Hill Book Company (UK) Limited, 4th edition, 1991.

[48] Afra Zomorodian and Gunnar Carlsson. Localized homology. *Proceedings of the IEEE International Conference on Shape Modeling and Applications 2007*, pages 189–198, 2007.

[49] Afra J. Zomorodian. *Topology for Computing*. Cambridge University Press, 1st edition, 2005.

# Appendix A

# Additional table and figures

This appendix contains extra tables and figures, referenced throughout the main body of the text.

| | Description | Dental=1 Skeletal=2 |
|---|---|---|
| Landmark 1 | Foramen Spinosum Left | 1 |
| Landmark 2 | Foramen Spinosum Right | 1 |
| Landmark 3 | ELSA | 1 |
| Landmark 4 | Auditory External Meatus Left | 1 |
| Landmark 5 | Auditory External Meatus Right | 1 |
| Landmark 6 | Dorsum Foramen Magnum | 1 |
| Landmark 7 | N | 1 |
| Landmark 8 | A | 1 |
| Landmark 9 | B | 1 |
| Landmark 10 | Prosthion | 1 |
| Landmark 11 | Mesial Incisor Surface Left | 2 |
| Landmark 12 | Mesial Incisor Surface Right | 2 |
| Landmark 13 | Zm Left | 1 |
| Landmark 14 | Zm Right | 1 |
| Landmark 15 | Piriform Left | 1 |
| Landmark 16 | Piriform Right | 1 |
| Landmark 17 | Orbit Left | 1 |
| Landmark 18 | Orbit Right | 1 |
| Landmark 19 | Ekm Left | 1 |
| Landmark 20 | Upper First Molar Left | 2 |
| Landmark 21 | Lower First Molar Left | 2 |
| Landmark 22 | Upper First PreMolar Left | 2 |
| Landmark 23 | Upper Canine Left | 2 |
| Landmark 24 | Lower Canine Left | 2 |
| Landmark 25 | Ekm Right | 1 |
| Landmark 26 | Upper First Molar Right | 2 |
| Landmark 27 | Lower First Molar Right | 2 |
| Landmark 28 | Upper First Premolar Right | 2 |
| Landmark 29 | Upper Canine Right | 2 |
| Landmark 30 | Lower Canine Right | 2 |
| Landmark 31 | Incisal Apex Left | 2 |
| Landmark 32 | Incisal Apex Right | 2 |
| Landmark 33 | MB 26 Apex | 2 |
| Landmark 34 | B 24 Apex | 2 |
| Landmark 35 | 23 Apex | 2 |
| Landmark 36 | MB 16 Apex | 2 |
| Landmark 37 | B 14 Apex | 2 |
| Landmark 38 | 13 Apex | 2 |
| Landmark 39 | MB 36 Apex | 2 |
| Landmark 40 | 33 Apex | 2 |
| Landmark 41 | MB 46 Apex | 2 |
| Landmark 42 | 43 Apex | 2 |
| Landmark 43 | ANS | 1 |
| Landmark 44 | PNS | 1 |

Table A.1: Description of Landmarks

| | x | y | z |
|---|---|---|---|
| Landmark 1 | 0.9958084 | 0.9972786 | 0.9948324 |
| Landmark 2 | 0.9958753 | 0.9987695 | 0.9980485 |
| Landmark 3 | 0.9961253 | 0.9983444 | 0.996648 |
| Landmark 4 | 0.965559 | 0.9968548 | 0.9970302 |
| Landmark 5 | 0.9819458 | 0.9948078 | 0.9971991 |
| Landmark 6 | 0.99657 | 0.9948347 | 0.9623226 |
| Landmark 7 | 0.997517 | 0.999795 | 0.9967474 |
| Landmark 8 | 0.9928113 | 0.9986231 | 0.9829405 |
| Landmark 9 | 0.9941482 | 0.999415 | 0.990828 |
| Landmark 10 | 0.9989228 | 0.9987176 | 0.9991659 |
| Landmark 11 | 0.9967599 | 0.9988233 | 0.9975326 |
| Landmark 12 | 0.9972109 | 0.9988184 | 0.9975326 |
| Landmark 13 | 0.9628446 | 0.9911847 | 0.9949996 |
| Landmark 14 | 0.9690635 | 0.989862 | 0.9961805 |
| Landmark 15 | 0.9992411 | 0.998466 | 0.9769627 |
| Landmark 16 | 0.997794 | 0.9979222 | 0.9775703 |
| Landmark 17 | 0.9890891 | 0.9973361 | 0.998431 |
| Landmark 18 | 0.9890126 | 0.9981342 | 0.9991872 |
| Landmark 19 | 0.9959578 | 0.996977 | 0.9841908 |
| Landmark 20 | 0.9993784 | 0.9974448 | 0.9979402 |
| Landmark 21 | 0.9971026 | 0.9991477 | 0.998717 |
| Landmark 22 | 0.996957 | 0.9971722 | 0.9965208 |
| Landmark 23 | 0.9973607 | 0.9993017 | 0.997304 |
| Landmark 24 | 0.998302 | 0.9989519 | 0.996325 |
| Landmark 25 | 0.9945508 | 0.9930129 | 0.976458 |
| Landmark 26 | 0.998807 | 0.998055 | 0.9984061 |
| Landmark 27 | 0.9971665 | 0.9981702 | 0.9980182 |
| Landmark 28 | 0.9975601 | 0.9972354 | 0.9974624 |
| Landmark 29 | 0.9972975 | 0.9984664 | 0.996562 |
| Landmark 30 | 0.9965226 | 0.9984577 | 0.9955306 |
| Landmark 31 | 0.9965413 | 0.9982377 | 0.9957652 |
| Landmark 32 | 0.9927915 | 0.9983424 | 0.9960272 |
| Landmark 33 | 0.9955616 | 0.9984745 | 0.9890536 |
| Landmark 34 | 0.997223 | 0.9987528 | 0.9895584 |
| Landmark 35 | 0.995864 | 0.9985045 | 0.9941512 |
| Landmark 36 | 0.9961252 | 0.998128 | 0.993571 |
| Landmark 37 | 0.9955913 | 0.9982972 | 0.9911 |
| Landmark 38 | 0.9955737 | 0.9984442 | 0.9954153 |
| Landmark 39 | 0.9969127 | 0.9970679 | 0.9874707 |
| Landmark 40 | 0.9962693 | 0.9987481 | 0.9900583 |
| Landmark 41 | 0.9981456 | 0.9977977 | 0.9943247 |
| Landmark 42 | 0.9958057 | 0.998944 | 0.9931568 |
| Landmark 43 | 0.9975644 | 0.9954345 | 0.9953342 |
| Landmark 44 | 0.9959843 | 0.9953744 | 0.9956113 |

Table A.2: ICCs for intra-rater reliability on each landmark

| | x | y | z |
|---|---|---|---|
| Landmark 1 | 0.9913705 | 0.992978 | 0.991311 |
| Landmark 2 | 0.9709572 | 0.9880345 | 0.9917325 |
| Landmark 3 | 0.9845[illegible]5 | 0.9940028 | 0.9924118 |
| Landmark 4 | 0.7656332 | 0.9805451 | 0.9882663 |
| Landmark 5 | 0.8739365 | 0.9837442 | 0.990004 |
| Landmark 6 | 0.9959398 | 0.9848262 | 0.935523 |
| Landmark 7 | 0.9968808 | 0.999529 | 0.9849608 |
| Landmark 8 | 0.9881966 | 0.9961152 | 0.9723194 |
| Landmark 9 | 0.9888079 | 0.9988344 | 0.983679 |
| Landmark 10 | 0.9976383 | 0.9973835 | 0.994522 |
| Landmark 11 | 0.994562 | 0.9952576 | 0.9910677 |
| Landmark 12 | 0.994562 | 0.9952576 | 0.9910677 |
| Landmark 13 | 0.9611155 | 0.9911701 | 0.9930684 |
| Landmark 14 | 0.9216593 | 0.9875174 | 0.9957658 |
| Landmark 15 | 0.9886363 | 0.9662529 | 0.914469 |
| Landmark 16 | 0.9885143 | 0.9840966 | 0.9034827 |
| Landmark 17 | 0.8830715 | 0.9911683 | 0.9975495 |
| Landmark 18 | 0.8686925 | 0.9863763 | 0.9986416 |
| Landmark 19 | 0.97994 | 0.9893867 | 0.9683173 |
| Landmark 20 | 0.9917479 | 0.9941226 | 0.996503 |
| Landmark 21 | 0.9940272 | 0.998107 | 0.9964064 |
| Landmark 22 | 0.9928393 | 0.990876 | 0.997138 |
| Landmark 23 | 0.9929314 | 0.998113 | 0.9923879 |
| Landmark 24 | 0.9881252 | 0.9951093 | 0.9962004 |
| Landmark 25 | 0.988708 | 0.9875736 | 0.963521 |
| Landmark 26 | 0.9929599 | 0.9973233 | 0.9960153 |
| Landmark 27 | 0.9962314 | 0.997107 | 0.9961063 |
| Landmark 28 | 0.9874819 | 0.9961817 | 0.9954411 |
| Landmark 29 | 0.9955638 | 0.9984212 | 0.9932852 |
| Landmark 30 | 0.9976362 | 0.9986385 | 0.9945674 |
| Landmark 31 | 0.996072 | 0.9973316 | 0.9942249 |
| Landmark 32 | 0.9951945 | 0.9962598 | 0.9927709 |
| Landmark 33 | 0.9958129 | 0.9978357 | 0.9851345 |
| Landmark 34 | 0.9934761 | 0.9987754 | 0.9866418 |
| Landmark 35 | 0.9932356 | 0.9971635 | 0.9953396 |
| Landmark 36 | 0.9911672 | 0.9960603 | 0.9875182 |
| Landmark 37 | 0.9898706 | 0.998135 | 0.9894108 |
| Landmark 38 | 0.9953716 | 0.997049 | 0.9959521 |
| Landmark 39 | 0.9917923 | 0.987671 | 0.965574 |
| Landmark 40 | 0.9935137 | 0.9974785 | 0.984006 |
| Landmark 41 | 0.9966715 | 0.9936912 | 0.9705181 |
| Landmark 42 | 0.9953077 | 0.9979976 | 0.9876961 |
| Landmark 43 | 0.9955007 | 0.9793602 | 0.980045 |
| Landmark 44 | 0.9917739 | 0.9845914 | 0.9916456 |

Table A.3: ICCs for inter-rater reliability on each landmark

| | Description |
|---|---|
| Landmark 1 | Foramen Spinosum Left |
| Landmark 2 | Foramen Spinosum Right |
| Landmark 3 | Ovale Left |
| Landmark 4 | Ovale Right |
| Landmark 5 | Hypoglossal Canal Left |
| Landmark 6 | Hypoglossal Canal Right |
| Landmark 7 | Rotundum Left |
| Landmark 8 | Rotundum Right |

Table A.4: Skull foramina chosen as landmarks.

| | Manuel1 | | | Manuel2 | | | Manuel3 | | | Jill | | | Carlos | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | x | y | z | x | y | z | x | y | z | x | y | z | x | y | z |
| Landmark 1 | 0.789 | 0.738 | 0.766 | 0.797 | 0.704 | 0.773 | 0.797 | 0.74 | 0.773 | 0.801 | 0.709 | 0.758 | 0.806 | 0.737 | 0.767 |
| Landmark 2 | 0.79 | 0.522 | 0.718 | 0.779 | 0.556 | 0.71 | 0.769 | 0.562 | 0.723 | 0.793 | 0.542 | 0.69 | 0.79 | 0.541 | 0.685 |
| Landmark 3 | 0.722 | 0.69 | 0.631 | 0.723 | 0.697 | 0.646 | 0.71 | 0.686 | 0.641 | 0.733 | 0.684 | 0.638 | 0.71 | 0.677 | 0.612 |
| Landmark 4 | 0.666 | 0.608 | 0.631 | 0.681 | 0.609 | 0.638 | 0.654 | 0.612 | 0.63 | 0.684 | 0.632 | 0.641 | 0.685 | 0.608 | 0.631 |
| Landmark 5 | 0.804 | 0.825 | 0.807 | 0.823 | 0.847 | 0.806 | 0.82 | 0.813 | 0.826 | 0.823 | 0.826 | 0.818 | 0.816 | 0.76 | 0.812 |
| Landmark 6 | 0.877 | 0.717 | 0.784 | 0.864 | 0.738 | 0.768 | 0.865 | 0.71 | 0.787 | 0.864 | 0.73 | 0.757 | 0.896 | 0.669 | 0.758 |
| Landmark 7 | 0.625 | 0.695 | 0.552 | 0.632 | 0.708 | 0.55 | 0.641 | 0.76 | 0.572 | 0.639 | 0.706 | 0.51 | 0.649 | 0.738 | 0.567 |
| Landmark 8 | 0.755 | 0.581 | 0.479 | 0.677 | 0.644 | 0.49 | 0.687 | 0.653 | 0.495 | 0.681 | 0.648 | 0.494 | 0.674 | 0.623 | 0.484 |

Table A.5: Mean measurement error by landmark coordinate for each rater, as compared to gold standard.

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| group | 2 | 0.000382 | 0.000191 | 0.1542 | 0.8574 |
| Residuals | 59 | 0.073098 | 0.001239 | | |

Table A.6: ANOVA table using first principal component of tangent space coordinates to compare treatment groups at baseline.

| | Df | Pillai | approx F | num Df | den Df | Pr(>F) |
|---|---|---|---|---|---|---|
| group | 2 | 0.15977 | 0.97238 | 10 | 112 | 0.4715 |
| Residuals | 59 | | | | | |

Table A.7: MANOVA table using first five principal components of tangent space coordinates to compare treatment groups at baseline.

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| group | 2 | 0.000389 | 0.000195 | 0.16 | 0.8525 |
| Residuals | 59 | 0.071778 | 0.001217 | | |

Table A.8: ANOVA table using first principal component of tangent space coordinates to compare treatment groups at time point 2.

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| group | 2 | 0.000138 | 0.000069 | 0.0628 | 0.9392 |
| Residuals | 59 | 0.064834 | 0.001099 | | |

Table A.9: ANOVA table using first principal component of tangent space coordinates to compare treatment groups at time point 3.

| | Df | Sum Sq | Mean Sq | F value | Pr(>F) |
|---|---|---|---|---|---|
| group | 2 | 0.000025 | 0.000013 | 0.0127 | 0.9874 |
| Residuals | 59 | 0.058737 | 0.000996 | | |

Table A.10: ANOVA table using first principal component of tangent space coordinates to compare treatment groups at time point 4.

| | Df | Pillai | approx F | num Df | den Df | Pr(>F) |
|---|---|---|---|---|---|---|
| group | 2 | 0.26046 | 1.67698 | 10 | 112 | 0.09476 |
| Residuals | 59 | | | | | |

Table A.11: MANOVA table using first five principal components of tangent space coordinates to compare treatment groups at time point 2.

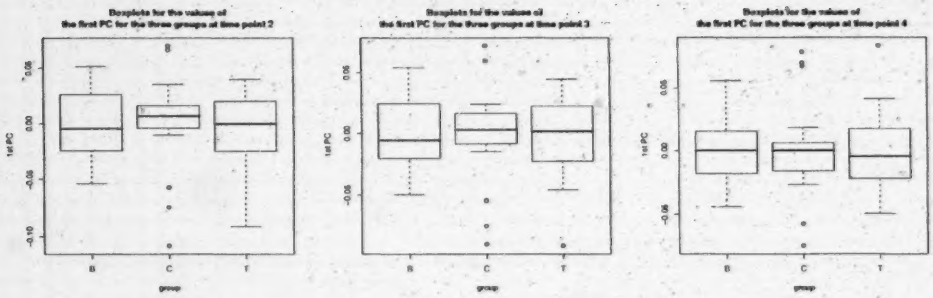

Figure A.1: Boxplots showing the distribution and variability of the three groups at each of the three time points.

| | Df | Pillai | approx F | num Df | den Df | Pr(>F) |
|---|---|---|---|---|---|---|
| group | 2 | 0.15978 | 1.23732 | 8 | 114 | 0.2839 |
| Residuals | 59 | | | | | |

Table A.12: MANOVA table using first four principal components of tangent space coordinates to compare treatment groups at time point 3.

| | Df | Pillai | approx F | num Df | den Df | Pr(>F) |
|---|---|---|---|---|---|---|
| group | 2 | 0.17186 | 1.33965 | 8 | 114 | 0.2311 |
| Residuals | 59 | | | | | |

Table A.13: MANOVA table using first four principal components of tangent space coordinates to compare treatment groups at time point 4.

| | B2-B1 | B3-B1 | B4-B1 | B3-B2 | B4-B2 | B4-B3 |
|---|---|---|---|---|---|---|
| $T_{obs}$ | 1.476944 | 1.457055 | 1.449564 | 1.427163 | 1.48592 | 1.220171 |
| $T_{(-1)}$ | 1.479494 | 1.457838 | 1.449507 | 1.43364 | 1.491514 | 1.220134 |
| $T_{(-2)}$ | 1.477136 | 1.45819 | 1.450908 | 1.430724 | 1.48851 | 1.220607 |
| $T_{(-3)}$ | 1.477073 | 1.457263 | 1.45 | 1.42599 | 1.486586 | 1.220403 |
| $T_{(-4)}$ | 1.48073 | 1.459855 | 1.450046 | 1.428549 | 1.484973 | 1.220269 |
| $T_{(-5)}$ | 1.480553 | 1.462784 | 1.451952 | 1.425720 | 1.481296 | 1.221113 |
| $T_{(-6)}$ | 1.467710 | 1.460130 | 1.449537 | 1.420134 | 1.487661 | 1.219390 |
| $T_{(-7)}$ | 1.481793 | 1.457129 | 1.443982 | 1.432562 | 1.481108 | 1.223396 |
| $T_{(-8)}$ | 1.477859 | 1.456695 | 1.414693 | 1.427472 | 1.466632 | 1.211824 |
| $T_{(-9)}$ | 1.475201 | 1.456951 | 1.450891 | 1.426955 | 1.487138 | 1.21901 |
| $T_{(-10)}$ | 1.477063 | 1.456845 | 1.414255 | 1.427013 | 1.465439 | 1.212000 |
| $T_{(-13)}$ | 1.471721 | 1.455018 | 1.449821 | 1.406351 | 1.470604 | 1.220565 |
| $T_{(-14)}$ | 1.484383 | 1.456582 | 1.451374 | 1.585125 | 1.629979 | 1.220679 |
| $T_{(-15)}$ | 1.479616 | 1.456807 | 1.449125 | 1.429456 | 1.486196 | 1.220276 |
| $T_{(-16)}$ | 1.478691 | 1.457241 | 1.449525 | 1.422805 | 1.478672 | 1.219938 |
| $T_{(-17)}$ | 1.473608 | 1.457093 | 1.447355 | 1.416925 | 1.475335 | 1.221848 |
| $T_{(-18)}$ | 1.474166 | 1.453966 | 1.446218 | 1.428148 | 1.486639 | 1.220334 |
| $T_{(-19)}$ | 1.476222 | 1.456459 | 1.450032 | 1.418525 | 1.479985 | 1.220262 |
| $T_{(-20)}$ | 1.477479 | 1.456375 | 1.449349 | 1.417064 | 1.477102 | 1.220146 |
| $T_{(-21)}$ | 1.477729 | 1.455422 | 1.449532 | 1.416939 | 1.477433 | 1.220118 |
| $T_{(-22)}$ | 1.475482 | 1.456695 | 1.449158 | 1.416748 | 1.476794 | 1.219954 |
| $T_{(-23)}$ | 1.472756 | 1.457259 | 1.449216 | 1.417324 | 1.477399 | 1.219833 |
| $T_{(-24)}$ | 1.478306 | 1.456536 | 1.448997 | 1.421701 | 1.481564 | 1.2197 |
| $T_{(-25)}$ | 1.480431 | 1.449101 | 1.313817 | 1.392832 | 1.326442 | 1.220379 |
| $T_{(-26)}$ | 1.479428 | 1.457587 | 1.450542 | 1.444829 | 1.501122 | 1.220783 |
| $T_{(-27)}$ | 1.477729 | 1.456641 | 1.450739 | 1.421581 | 1.483226 | 1.219697 |
| $T_{(-28)}$ | 1.477732 | 1.457037 | 1.448520 | 1.418713 | 1.477427 | 1.219957 |
| $T_{(-29)}$ | 1.475983 | 1.456723 | 1.448714 | 1.419525 | 1.478243 | 1.219952 |
| $T_{(-30)}$ | 1.476026 | 1.456300 | 1.449674 | 1.421833 | 1.482289 | 1.219593 |
| $T_{(-31)}$ | 1.435683 | 1.358342 | 1.449892 | 1.432292 | 1.488514 | 1.220094 |
| $T_{(-32)}$ | 1.301087 | 1.358001 | 1.449651 | 1.425429 | 1.482826 | 1.220167 |
| $T_{(-33)}$ | 1.475988 | 1.456726 | 1.448718 | 1.421701 | 1.481113 | 1.220024 |
| $T_{(-34)}$ | 1.474366 | 1.457422 | 1.449756 | 1.422499 | 1.482268 | 1.219937 |
| $T_{(-35)}$ | 1.470529 | 1.457282 | 1.449673 | 1.425245 | 1.484287 | 1.220344 |
| $T_{(-36)}$ | 1.478780 | 1.449431 | 1.312455 | 1.379062 | 1.324626 | 1.220428 |
| $T_{(-37)}$ | 1.438448 | 1.457827 | 1.449876 | 1.418734 | 1.477702 | 1.208213 |
| $T_{(-38)}$ | 1.440403 | 1.457135 | 1.449516 | 1.374293 | 1.477643 | 1.208244 |
| $T_{(-39)}$ | 1.479876 | 1.455324 | 1.449867 | 1.420540 | 1.480419 | 1.220378 |
| $T_{(-40)}$ | 1.478730 | 1.456822 | 1.450018 | 1.425908 | 1.485241 | 1.219513 |
| $T_{(-41)}$ | 1.472248 | 1.456465 | 1.450141 | 1.412969 | 1.475930 | 1.218718 |
| $T_{(-42)}$ | 1.474264 | 1.456877 | 1.451346 | 1.422805 | 1.485591 | 1.219301 |
| $T_{(-43)}$ | 1.479135 | 1.456902 | 1.45034 | 1.427964 | 1.486376 | 1.219949 |
| $T_{(-44)}$ | 1.476716 | 1.45716 | 1.449870 | 1.426227 | 1.486244 | 1.220187 |

Table A.14: Observed and landmark-deleted T-statistics when comparing treatment B groups across time.

| | C3-C1 | C4-C1 | C4-C3 |
|---|---|---|---|
| $T_{obs}$ | 1.315932 | 1.370451 | 1.175318 |
| $T_{(-1)}$ | 1.319288 | 1.371343 | 1.176466 |
| $T_{(-2)}$ | 1.316406 | 1.369777 | 1.173421 |
| $T_{(-3)}$ | 1.316104 | 1.370586 | 1.175563 |
| $T_{(-4)}$ | 1.319307 | 1.372827 | 1.176538 |
| $T_{(-5)}$ | 1.321336 | 1.370006 | 1.173402 |
| $T_{(-6)}$ | 1.316147 | 1.371104 | 1.171375 |
| $T_{(-7)}$ | 1.315184 | 1.372627 | 1.172536 |
| $T_{(-8)}$ | 1.315901 | 1.370303 | 1.146902 |
| $T_{(-9)}$ | 1.315668 | 1.368744 | 1.176025 |
| $T_{(-10)}$ | 1.316028 | 1.37064 | 1.175534 |
| $T_{(-13)}$ | 1.315730 | 1.366281 | 1.174443 |
| $T_{(-14)}$ | 1.293815 | 1.372774 | 1.175661 |
| $T_{(-15)}$ | 1.316061 | 1.370384 | 1.176048 |
| $T_{(-16)}$ | 1.315527 | 1.370701 | 1.174566 |
| $T_{(-17)}$ | 1.316996 | 1.370270 | 1.174734 |
| $T_{(-18)}$ | 1.316037 | 1.368887 | 1.175334 |
| $T_{(-19)}$ | 1.316453 | 1.369545 | 1.175467 |
| $T_{(-20)}$ | 1.314307 | 1.369779 | 1.17531 |
| $T_{(-21)}$ | 1.313841 | 1.369345 | 1.174626 |
| $T_{(-22)}$ | 1.315196 | 1.370445 | 1.175303 |
| $T_{(-23)}$ | 1.314228 | 1.370278 | 1.175575 |
| $T_{(-24)}$ | 1.315487 | 1.370232 | 1.175938 |
| $T_{(-25)}$ | 1.197812 | 1.209733 | 1.176598 |
| $T_{(-26)}$ | 1.316064 | 1.374116 | 1.176408 |
| $T_{(-27)}$ | 1.31159 | 1.368954 | 1.175602 |
| $T_{(-28)}$ | 1.317077 | 1.372292 | 1.175209 |
| $T_{(-29)}$ | 1.317398 | 1.371780 | 1.174939 |
| $T_{(-30)}$ | 1.314208 | 1.367036 | 1.175078 |
| $T_{(-31)}$ | 1.315902 | 1.370367 | 1.159399 |
| $T_{(-32)}$ | 1.315996 | 1.370389 | 1.175397 |
| $T_{(-33)}$ | 1.316472 | 1.370078 | 1.175194 |
| $T_{(-34)}$ | 1.316243 | 1.370219 | 1.175238 |
| $T_{(-35)}$ | 1.316020 | 1.370576 | 1.175299 |
| $T_{(-36)}$ | 1.217474 | 1.210009 | 1.148749 |
| $T_{(-37)}$ | 1.316818 | 1.370029 | 1.148303 |
| $T_{(-38)}$ | 1.315452 | 1.369119 | 1.17585 |
| $T_{(-39)}$ | 1.314778 | 1.369086 | 1.174630 |
| $T_{(-40)}$ | 1.315244 | 1.370396 | 1.174700 |
| $T_{(-41)}$ | 1.314875 | 1.372008 | 1.176034 |
| $T_{(-42)}$ | 1.314750 | 1.368458 | 1.175452 |
| $T_{(-43)}$ | 1.315810 | 1.370354 | 1.176154 |
| $T_{(-44)}$ | 1.315877 | 1.370197 | 1.175613 |

Table A.15: Observed and landmark-deleted T-statistics when comparing treatment C groups across time.

| | T2-T1 | T3-T1 | T4-T1 | T3-T2 | T4-T2 | T4-T3 |
|---|---|---|---|---|---|---|
| $T_{obs}$ | 1.663264 | 1.624671 | 1.702729 | 1.423590 | 1.591574 | 1.304795 |
| $T_{(-1)}$ | 1.663564 | 1.624195 | 1.703715 | 1.423314 | 1.59277 | 1.304539 |
| $T_{(-2)}$ | 1.663012 | 1.623673 | 1.704582 | 1.422830 | 1.593378 | 1.304717 |
| $T_{(-3)}$ | 1.663258 | 1.624696 | 1.703256 | 1.423406 | 1.592030 | 1.304451 |
| $T_{(-4)}$ | 1.662931 | 1.624536 | 1.702147 | 1.421203 | 1.588234 | 1.305092 |
| $T_{(-5)}$ | 1.664150 | 1.621448 | 1.704732 | 1.420521 | 1.588869 | 1.3049 |
| $T_{(-6)}$ | 1.661742 | 1.627725 | 1.697096 | 1.423236 | 1.592664 | 1.306080 |
| $T_{(-7)}$ | 1.668905 | 1.630091 | 1.701936 | 1.423388 | 1.58827 | 1.305385 |
| $T_{(-8)}$ | 1.650123 | 1.623937 | 1.580424 | 1.423320 | 1.507507 | 1.209286 |
| $T_{(-9)}$ | 1.663151 | 1.623993 | 1.703906 | 1.423964 | 1.592032 | 1.305426 |
| $T_{(-10)}$ | 1.663188 | 1.624417 | 1.703597 | 1.423437 | 1.507485 | 1.247401 |
| $T_{(-13)}$ | 1.661376 | 1.623585 | 1.704096 | 1.416443 | 1.591945 | 1.305008 |
| $T_{(-14)}$ | 1.663097 | 1.624629 | 1.704747 | 1.423406 | 1.592182 | 1.304793 |
| $T_{(-15)}$ | 1.665319 | 1.62576 | 1.700130 | 1.424400 | 1.591015 | 1.302675 |
| $T_{(-16)}$ | 1.664173 | 1.624680 | 1.700290 | 1.423495 | 1.590952 | 1.300663 |
| $T_{(-17)}$ | 1.663846 | 1.631803 | 1.701433 | 1.425049 | 1.589149 | 1.304183 |
| $T_{(-18)}$ | 1.662092 | 1.623453 | 1.700110 | 1.421666 | 1.590211 | 1.303259 |
| $T_{(-19)}$ | 1.663261 | 1.513680 | 1.703215 | 1.378578 | 1.591342 | 1.304853 |
| $T_{(-20)}$ | 1.662948 | 1.624173 | 1.703345 | 1.421176 | 1.592672 | 1.304364 |
| $T_{(-21)}$ | 1.661902 | 1.626045 | 1.702299 | 1.423690 | 1.592045 | 1.304741 |
| $T_{(-22)}$ | 1.663315 | 1.624523 | 1.702328 | 1.425500 | 1.591736 | 1.304781 |
| $T_{(-23)}$ | 1.663337 | 1.623997 | 1.702519 | 1.423890 | 1.592008 | 1.304725 |
| $T_{(-24)}$ | 1.663843 | 1.624350 | 1.702651 | 1.423910 | 1.591371 | 1.305706 |
| $T_{(-25)}$ | 1.662725 | 1.624619 | 1.590201 | 1.423206 | 1.523007 | 1.304862 |
| $T_{(-26)}$ | 1.663215 | 1.625312 | 1.702671 | 1.424976 | 1.591740 | 1.304425 |
| $T_{(-27)}$ | 1.662427 | 1.623111 | 1.701853 | 1.424779 | 1.5915 | 1.304185 |
| $T_{(-28)}$ | 1.663372 | 1.625234 | 1.702938 | 1.42431 | 1.592292 | 1.304692 |
| $T_{(-29)}$ | 1.663007 | 1.625887 | 1.704320 | 1.425162 | 1.592826 | 1.304649 |
| $T_{(-30)}$ | 1.66372 | 1.624441 | 1.703968 | 1.423862 | 1.592689 | 1.305072 |
| $T_{(-31)}$ | 1.503865 | 1.508641 | 1.703817 | 1.423501 | 1.591706 | 1.305084 |
| $T_{(-32)}$ | 1.285437 | 1.479377 | 1.580487 | 1.423633 | 1.591538 | 1.305201 |
| $T_{(-33)}$ | 1.662959 | 1.513523 | 1.704218 | 1.378657 | 1.593073 | 1.304879 |
| $T_{(-34)}$ | 1.66327 | 1.623522 | 1.704259 | 1.423817 | 1.59289 | 1.304853 |
| $T_{(-35)}$ | 1.663146 | 1.624458 | 1.703555 | 1.422085 | 1.592747 | 1.305037 |
| $T_{(-36)}$ | 1.662910 | 1.625312 | 1.593148 | 1.423898 | 1.523069 | 1.304885 |
| $T_{(-37)}$ | 1.663231 | 1.625180 | 1.702135 | 1.423580 | 1.591278 | 1.30493 |
| $T_{(-38)}$ | 1.663751 | 1.624979 | 1.702119 | 1.423565 | 1.590859 | 1.305529 |
| $T_{(-39)}$ | 1.661956 | 1.625607 | 1.702349 | 1.426281 | 1.593238 | 1.304681 |
| $T_{(-40)}$ | 1.664467 | 1.624501 | 1.704638 | 1.423499 | 1.592077 | 1.308928 |
| $T_{(-41)}$ | 1.661683 | 1.622299 | 1.699575 | 1.426083 | 1.58841 | 1.304130 |
| $T_{(-42)}$ | 1.664271 | 1.622897 | 1.703151 | 1.424429 | 1.590632 | 1.306598 |
| $T_{(-43)}$ | 1.644251 | 1.623714 | 1.702411 | 1.423263 | 1.592178 | 1.268919 |
| $T_{(-44)}$ | 1.662956 | 1.624665 | 1.702059 | 1.423668 | 1.591724 | 1.304615 |

Table A.16: Observed and landmark-deleted T-statistics when comparing treatment T groups across time.

| | B2 and C2 | B2 and T2 | T2 and B2 | T2 and C2 |
|---|---|---|---|---|
| $T_{obs}$ | 1.418469 | 1.327263 | 1.327263 | 1.498158 |
| $T_{(-1)}$ | 1.421057 | 1.327151 | 1.327151 | 1.499806 |
| $T_{(-2)}$ | 1.418844 | 1.327312 | 1.327312 | 1.499438 |
| $T_{(-3)}$ | 1.417677 | 1.327010 | 1.327010 | 1.499799 |
| $T_{(-4)}$ | 1.422747 | 1.327007 | 1.327007 | 1.498643 |
| $T_{(-5)}$ | 1.422644 | 1.328607 | 1.328607 | 1.498470 |
| $T_{(-6)}$ | 1.423133 | 1.326624 | 1.326624 | 1.498739 |
| $T_{(-7)}$ | 1.419444 | 1.327639 | 1.327639 | 1.499264 |
| $T_{(-8)}$ | 1.418682 | 1.327265 | 1.327265 | 1.498187 |
| $T_{(-9)}$ | 1.418538 | 1.326426 | 1.326426 | 1.497293 |
| $T_{(-10)}$ | 1.418672 | 1.327150 | 1.327150 | 1.498234 |
| $T_{(-13)}$ | 1.416068 | 1.327374 | 1.327374 | 1.497706 |
| $T_{(-14)}$ | 1.421096 | 1.327941 | 1.327941 | 1.498172 |
| $T_{(-15)}$ | 1.417625 | 1.32671 | 1.32671 | 1.497900 |
| $T_{(-16)}$ | 1.417305 | 1.328517 | 1.328517 | 1.499008 |
| $T_{(-17)}$ | 1.415780 | 1.326895 | 1.326895 | 1.499071 |
| $T_{(-18)}$ | 1.416742 | 1.329412 | 1.329412 | 1.497295 |
| $T_{(-19)}$ | 1.419626 | 1.327266 | 1.327266 | 1.497649 |
| $T_{(-20)}$ | 1.419637 | 1.32714 | 1.32714 | 1.498525 |
| $T_{(-21)}$ | 1.418910 | 1.327126 | 1.327126 | 1.497898 |
| $T_{(-22)}$ | 1.417624 | 1.327059 | 1.327059 | 1.498421 |
| $T_{(-23)}$ | 1.415624 | 1.327464 | 1.327464 | 1.498401 |
| $T_{(-24)}$ | 1.418972 | 1.326671 | 1.326671 | 1.497536 |
| $T_{(-25)}$ | 1.418721 | 1.327530 | 1.327530 | 1.498144 |
| $T_{(-26)}$ | 1.419371 | 1.305505 | 1.305505 | 1.492794 |
| $T_{(-27)}$ | 1.41907 | 1.306661 | 1.306661 | 1.491854 |
| $T_{(-28)}$ | 1.417783 | 1.327646 | 1.327646 | 1.498118 |
| $T_{(-29)}$ | 1.418173 | 1.307336 | 1.307336 | 1.498392 |
| $T_{(-30)}$ | 1.418011 | 1.308014 | 1.308014 | 1.497593 |
| $T_{(-31)}$ | 1.354236 | 1.327260 | 1.327260 | 1.267227 |
| $T_{(-32)}$ | 1.355613 | 1.327220 | 1.327220 | 1.241104 |
| $T_{(-33)}$ | 1.419151 | 1.327259 | 1.327259 | 1.497974 |
| $T_{(-34)}$ | 1.416951 | 1.327265 | 1.327265 | 1.498368 |
| $T_{(-35)}$ | 1.412659 | 1.327211 | 1.327211 | 1.497841 |
| $T_{(-36)}$ | 1.417374 | 1.325165 | 1.325165 | 1.498329 |
| $T_{(-37)}$ | 1.418255 | 1.327402 | 1.327402 | 1.498109 |
| $T_{(-38)}$ | 1.418183 | 1.328463 | 1.328463 | 1.497493 |
| $T_{(-39)}$ | 1.420221 | 1.325932 | 1.325932 | 1.498942 |
| $T_{(-40)}$ | 1.419162 | 1.324437 | 1.324437 | 1.496321 |
| $T_{(-41)}$ | 1.419648 | 1.327006 | 1.327006 | 1.496325 |
| $T_{(-42)}$ | 1.418068 | 1.329836 | 1.329836 | 1.497532 |
| $T_{(-43)}$ | 1.419148 | 1.326427 | 1.326427 | 1.498129 |
| $T_{(-44)}$ | 1.418333 | 1.327272 | 1.327272 | 1.498320 |

Table A.17: Observed and landmark-deleted T-statistics when comparing treatment groups at time point 2.

| | B3 and C3 | B3 and T3 | T3 and B3 | T3 and C3 |
|---|---|---|---|---|
| $T_{obs}$ | 1.338226 | 1.350373 | 1.350373 | 1.391712 |
| $T_{(-1)}$ | 1.338087 | 1.350966 | 1.350966 | 1.392407 |
| $T_{(-2)}$ | 1.337862 | 1.347457 | 1.347457 | 1.389937 |
| $T_{(-3)}$ | 1.338310 | 1.350098 | 1.350098 | 1.390922 |
| $T_{(-4)}$ | 1.336446 | 1.348091 | 1.348091 | 1.390925 |
| $T_{(-5)}$ | 1.336689 | 1.348452 | 1.348452 | 1.387216 |
| $T_{(-6)}$ | 1.338947 | 1.347358 | 1.347358 | 1.393691 |
| $T_{(-7)}$ | 1.334334 | 1.35251 | 1.35251 | 1.398260 |
| $T_{(-8)}$ | 1.338181 | 1.350308 | 1.350308 | 1.390907 |
| $T_{(-9)}$ | 1.339640 | 1.350143 | 1.350143 | 1.392129 |
| $T_{(-10)}$ | 1.338269 | 1.350953 | 1.350953 | 1.391658 |
| $T_{(-13)}$ | 1.339417 | 1.347382 | 1.347382 | 1.385907 |
| $T_{(-14)}$ | 1.338149 | 1.351179 | 1.351179 | 1.392044 |
| $T_{(-15)}$ | 1.337021 | 1.352424 | 1.352424 | 1.392321 |
| $T_{(-16)}$ | 1.337365 | 1.351498 | 1.351498 | 1.392897 |
| $T_{(-17)}$ | 1.336787 | 1.352484 | 1.352484 | 1.392247 |
| $T_{(-18)}$ | 1.339414 | 1.349653 | 1.349653 | 1.388599 |
| $T_{(-19)}$ | 1.339082 | 1.324355 | 1.324355 | 1.284218 |
| $T_{(-20)}$ | 1.337468 | 1.349311 | 1.349311 | 1.389970 |
| $T_{(-21)}$ | 1.338963 | 1.350058 | 1.350058 | 1.392731 |
| $T_{(-22)}$ | 1.337731 | 1.350968 | 1.350968 | 1.391874 |
| $T_{(-23)}$ | 1.331477 | 1.350243 | 1.350243 | 1.39001 |
| $T_{(-24)}$ | 1.331095 | 1.351015 | 1.351015 | 1.394574 |
| $T_{(-25)}$ | 1.338652 | 1.351709 | 1.351709 | 1.393346 |
| $T_{(-26)}$ | 1.338014 | 1.351267 | 1.351267 | 1.3938 |
| $T_{(-27)}$ | 1.337134 | 1.351600 | 1.351600 | 1.391522 |
| $T_{(-28)}$ | 1.338380 | 1.294328 | 1.294328 | 1.337417 |
| $T_{(-29)}$ | 1.338273 | 1.290893 | 1.290893 | 1.337913 |
| $T_{(-30)}$ | 1.338017 | 1.352188 | 1.352188 | 1.392387 |
| $T_{(-31)}$ | 1.252268 | 1.350340 | 1.350340 | 1.391181 |
| $T_{(-32)}$ | 1.252216 | 1.350324 | 1.350324 | 1.391486 |
| $T_{(-33)}$ | 1.339414 | 1.317239 | 1.317239 | 1.284176 |
| $T_{(-34)}$ | 1.338355 | 1.349510 | 1.349510 | 1.392010 |
| $T_{(-35)}$ | 1.338518 | 1.348796 | 1.348796 | 1.390698 |
| $T_{(-36)}$ | 1.338601 | 1.349604 | 1.349604 | 1.39187 |
| $T_{(-37)}$ | 1.338404 | 1.350125 | 1.350125 | 1.391801 |
| $T_{(-38)}$ | 1.338891 | 1.349989 | 1.349989 | 1.391973 |
| $T_{(-39)}$ | 1.338932 | 1.349182 | 1.349182 | 1.390087 |
| $T_{(-40)}$ | 1.339043 | 1.349819 | 1.349819 | 1.392184 |
| $T_{(-41)}$ | 1.338001 | 1.351576 | 1.351576 | 1.394007 |
| $T_{(-42)}$ | 1.339981 | 1.352398 | 1.352398 | 1.392872 |
| $T_{(-43)}$ | 1.338363 | 1.350436 | 1.350436 | 1.390653 |
| $T_{(-44)}$ | 1.338247 | 1.350318 | 1.350318 | 1.392049 |

Table A.18: Observed and landmark-deleted T-statistics when comparing treatment groups at time point 3.

| | B4 and C4 | B4 and T4 | T4 and B4 | T4 and C4 |
|---|---|---|---|---|
| $T_{obs}$ | 1.28259 | 1.386536 | 1.386536 | 1.325425 |
| $T_{(-1)}$ | 1.282335 | 1.386541 | 1.386541 | 1.325604 |
| $T_{(-2)}$ | 1.283536 | 1.386924 | 1.386924 | 1.326779 |
| $T_{(-3)}$ | 1.282630 | 1.386237 | 1.386237 | 1.325167 |
| $T_{(-4)}$ | 1.282295 | 1.384563 | 1.384563 | 1.325302 |
| $T_{(-5)}$ | 1.280178 | 1.389934 | 1.389934 | 1.332315 |
| $T_{(-6)}$ | 1.282216 | 1.385433 | 1.385433 | 1.324421 |
| $T_{(-7)}$ | 1.280218 | 1.386949 | 1.386949 | 1.316782 |
| $T_{(-8)}$ | 1.282559 | 1.386481 | 1.386481 | 1.325588 |
| $T_{(-9)}$ | 1.2834 | 1.386258 | 1.386258 | 1.327344 |
| $T_{(-10)}$ | 1.282714 | 1.38679 | 1.38679 | 1.325184 |
| $T_{(-13)}$ | 1.283089 | 1.383810 | 1.383810 | 1.323132 |
| $T_{(-14)}$ | 1.281083 | 1.386676 | 1.386676 | 1.326800 |
| $T_{(-15)}$ | 1.282984 | 1.385175 | 1.385175 | 1.325310 |
| $T_{(-16)}$ | 1.282504 | 1.385765 | 1.385765 | 1.325360 |
| $T_{(-17)}$ | 1.284343 | 1.387393 | 1.387393 | 1.326368 |
| $T_{(-18)}$ | 1.282805 | 1.384329 | 1.384329 | 1.323699 |
| $T_{(-19)}$ | 1.283258 | 1.368605 | 1.368605 | 1.288411 |
| $T_{(-20)}$ | 1.282540 | 1.387235 | 1.387235 | 1.326808 |
| $T_{(-21)}$ | 1.281770 | 1.383384 | 1.383384 | 1.321427 |
| $T_{(-22)}$ | 1.282325 | 1.387927 | 1.387927 | 1.324969 |
| $T_{(-23)}$ | 1.282208 | 1.387008 | 1.387008 | 1.322663 |
| $T_{(-24)}$ | 1.281740 | 1.387105 | 1.387105 | 1.326045 |
| $T_{(-25)}$ | 1.283681 | 1.386059 | 1.386059 | 1.325078 |
| $T_{(-26)}$ | 1.282146 | 1.386645 | 1.386645 | 1.325326 |
| $T_{(-27)}$ | 1.282520 | 1.388119 | 1.388119 | 1.325851 |
| $T_{(-28)}$ | 1.282249 | 1.38718 | 1.38718 | 1.325572 |
| $T_{(-29)}$ | 1.224404 | 1.277849 | 1.277849 | 1.326530 |
| $T_{(-30)}$ | 1.225343 | 1.368638 | 1.368638 | 1.325322 |
| $T_{(-31)}$ | 1.212968 | 1.386357 | 1.386357 | 1.325508 |
| $T_{(-32)}$ | 1.265043 | 1.386428 | 1.386428 | 1.325691 |
| $T_{(-33)}$ | 1.282996 | 1.334834 | 1.334834 | 1.289277 |
| $T_{(-34)}$ | 1.282758 | 1.385937 | 1.385937 | 1.309414 |
| $T_{(-35)}$ | 1.282794 | 1.385637 | 1.385637 | 1.309810 |
| $T_{(-36)}$ | 1.283854 | 1.387260 | 1.387260 | 1.326211 |
| $T_{(-37)}$ | 1.283069 | 1.387593 | 1.387593 | 1.326356 |
| $T_{(-38)}$ | 1.283306 | 1.388413 | 1.388413 | 1.326226 |
| $T_{(-39)}$ | 1.281178 | 1.382309 | 1.382309 | 1.323243 |
| $T_{(-40)}$ | 1.281672 | 1.385683 | 1.385683 | 1.328346 |
| $T_{(-41)}$ | 1.283040 | 1.388559 | 1.388559 | 1.325760 |
| $T_{(-42)}$ | 1.283419 | 1.388738 | 1.388738 | 1.327054 |
| $T_{(-43)}$ | 1.282515 | 1.386335 | 1.386335 | 1.325152 |
| $T_{(-44)}$ | 1.282585 | 1.386461 | 1.386461 | 1.325426 |

Table A.19: Observed and landmark-deleted T-statistics when comparing treatment groups at time point 4.

# Appendix B

# Computer code

The code below was written in R by Jennifer Gamble to implement EDMA methods as described in [31].

Required packages:

- `utils`
- `fields` (to use rdist() function)
- `corpcor` (to allow use of smtools like sm2vec)
- `abind` (to allow binding of arrays)

For all functions (unless otherwise stated), X and Y are $k \times m \times n_i$ arrays of raw landmark coordinate data.

List of functions:

```
bootFDM(X,Y,b) # returns unsorted 1xb matrix
    for bootstrapped FDM (p.23-24)


bootT(X,Y,b) # bootstrap for T in EDMA-I test,
    returns b (number reps), bootstrapped FDM
    (raw and sorted), bootstrapped T stats, and
    ninety and ninety-five percent CIs for T
```

```
btsort(boot) # takes bootFDM output and returns
    sorted lxb matrix, b, and ninety and ninety-
    five percent confidence intervals for each
    FDM entry (i.e. row)

F(A) # returns vectorized form matrix l=k(k-1)/2 of
    landmark configuration A

FDM(X,Y) # returns exact form difference matrix
    for two kxm configurations X and Y

FDMhat(X,Y) # returns estimated form difference
    matrix

FMhat(X) # returns estimated mean form matrix

index(A) # returns  lx2 matrix where row i gives the
    two landmarks corresponding to F(A) entry i

Mhat(X) # returns mean form estimate from step 7
    of algorithm to estimate mean form matrix (p.22)

T(FDM) # returns Lele and Richtsmeier's test statistic
    for given form difference matrix

Tdelete(X,Y) # returns k-length vector whose ith
    entry is the T statistic recalculated with
    the ith landmark removed
```

Code for functions:

```
bootFDM <- function(X, Y, b) {
```

```
k = dim(X)[1]
l = k*(k-1)/2
n1 = dim(X)[3]
n2 = dim(Y)[3]
F1all = matrix(nrow=l, ncol=n1)
for(i in 1:n1) {
 F1all[,i] = F(X[,,i])
    }
F2all = matrix(nrow=l, ncol=n2)
for(i in 1:n2) {
    F2all[,i] = F(Y[,,i])
    }
boot = matrix(nrow=1, ncol=b)
boot[,1] = FMhat(F1all)/FMhat(F2all)
for(j in 2:b) {
    F1boot = matrix(nrow=l, ncol=n1)
    rdm1 = sample(1:n1, n1, replace=TRUE)
    for(i in 1:n1) {
        F1boot[,i] = F1all[,rdm1[i]]
        }
    FM1 = FMhat(F1boot)
    F2boot = matrix(nrow=l, ncol=n2)
    rdm2 = sample(1:n2, n2, replace=TRUE)
    for(i in 1:n2) {
        F2boot[,i] = F2all[,rdm2[i]]
        }
    FM2 = FMhat(F2boot)
    boot[,j] = FM1/FM2
    }
boot
}
```

```
bootT <- function(X, Y, b) {
    k = dim(X)[1]
    l = k*(k-1)/2
    n1 = dim(X)[3]
    n2 = dim(Y)[3]
    F1all = matrix(nrow=l, ncol=n1)
    for(i in 1:n1) {
        F1all[,i] = F(X[,,i])
        }
    F2all = matrix(nrow=l, ncol=n2)
    for(i in 1:n2) {
        F2all[,i] = F(Y[,,i])
        }
    boot = matrix(nrow=l, ncol=b)
    T = vector(length=b)
    for(j in 1:b) {
        rdm1 = sample(1:n2, n1, replace=TRUE)
        Aall = matrix(nrow=l, ncol=n1)
        for(i in 1:n1) {
            Aall[,i] = F2all[,rdm1[i]]
            }
        rdm2 = sample(1:n2, n2, replace=TRUE)
        Ball = matrix(nrow=l, ncol=n2)
        for(i in 1:n2) {
            Ball[,i] = F2all[,rdm2[i]]
            }
        formdiff = FMhat(Ball)/FMhat(Aall)
        boot[,j] = formdiff
        T[j] = T(formdiff)
        }
```

```
    low95 = floor((b*0.025)+1)
    high95 = floor(b*0.975)
    low90 = floor((b*0.05)+1)
    high90 = floor(b*0.95)
    Tsort = sort(T)
    ninetyfive = c(Tsort[low95], Tsort[high95])
    ninety = c(Tsort[low90], Tsort[high90])
    return(list(b=b, boot=boot, T=T, ninetyfive=ninetyfive,
           ninety=ninety, Tsort=Tsort))
    }

btsort <- function(boot) {
    b = dim(boot)[2]
    bootsort = matrix(nrow=1, ncol=b)
    for(i in 1:1) {
        bootsort[i,] = sort(boot[i,])
        }
    low95 = floor((b*0.025)+1)
    high95 = floor(b*0.975)
    low90 = floor((b*0.05)+1)
    high90 = floor(b*0.95)
    ninetyfive = cbind(bootsort[,low95], bootsort[,high95],
                 index)
    ninety = cbind(bootsort[,low90], bootsort[,high90],
                 index)
    return(list(bootsort=bootsort, b=b,
        ninetyfive=ninetyfive, ninety=ninety))
    }

F <- function(A) sm2vec(rdist(A,A), diag=FALSE)
```

```
FDM <- function(X,Y) F(X)/F(Y)

FDMhat <- function(X, Y) {
    k = dim(X)[1]
    l = k*(k-1)/2
    n1 = dim(X)[3]
    n2 = dim(Y)[3]
    F.X = matrix(nrow=l, ncol=n1)
    for(i in 1:n1) {
        F.X[,i] = F(X[,,i])
        }
    F.Y = matrix(nrow=l, ncol=n2)
    for(i in 1:n2) {
        F.Y[,i] = F(Y[,,i])
        }
    FDM.XY = FMhat(F.X)/FMhat(F.Y)
    FDM.XY
    }

FMhat <- function(Fall) F(Mhat(Fall))

index <- function(A) sm.index(rdist(X(1),X(1)), diag=FALSE)

Mhat <- function(Fall) {
    k = dim(X)[1]
    l = k*(k-1)/2
    n1 = dim(X)[3]
    Fall = matrix(nrow=l, ncol=n1)
    for(i in 1:n1) {
        Fall[,i] = F(X[,,i])
        }
```

```
    Eall = Fall^2
    Ebar = rowMeans(Eall)
    var.E = vector(length=1)
    for(i in 1:1) var.E[i] = var(Eall[i,])
    EpsHat = ((Ebar^2)-1.5*var.E)^.5
    EM = vec2sm(EpsHat, diag=FALSE)
    k = dim(EM)[1]
    for(i in 1:k) EM[i,i]=0
    for(i in 1:k) {
        for(j in 1:k) {
            if(is.nan(EM)[i,j]==TRUE) EM[i,j]=0
            }
        }
    H = diag(k)-(1/k)
    B = (-.5)*H%*%EM%*%t(H)
    eval = eigen(B)$values
    evec = eigen(B)$vectors
    Mhat = cbind(sqrt(eval[1])*evec[,1], sqrt(eval[2])*evec[,2],
        sqrt(eval[3])*evec[,3])
    Mhat
    }

T <- function(FDM) max(FDM)/min(FDM)

Tdelete <- function(X, Y) {
    Tdelete = vector(mode="numeric", length=dim(X)[1])
    for(i in 1:dim(X)[1]) {
        Xi = X[-i,,]
        Yi = Y[-i,,]
        FDMi = FDMhat(Xi,Yi)
        Ti = T(FDMi)
```

```
    Tdelete[i] = Ti
    }
Tdelete
}
```